

El Periodismo de Datos como herramienta para generar noticias e investigaciones



Eugenia Michiels.

Licenciatura en Periodismo.
Universidad Abierta Interamericana.
Agosto 2017.

Resumen

Actualmente el panorama de los medios de comunicación a nivel mundial tiene una nueva configuración dada por la generalización de Internet y su influencia como principal fuente de información para el trabajo del periodista. En este contexto y con la implementación de nuevas tecnologías de la información, ha surgido una nueva disciplina en auge: el Periodismo de Datos.

El presente trabajo de investigación pretende realizar una descripción cualitativa de la disciplina denominada Periodismo de Datos, considerando sus aspectos más importantes y su aplicación como herramienta para llevar a cabo noticias e investigaciones.

Centrado en el análisis de las dos principales entidades de la ciudad de Rosario que trabajan con esta disciplina: el diario digital Porven y la organización Hacks/Hackers Rosario; mediante la observación de sus metodologías de trabajo y sus investigaciones realizadas en materia de Periodismo de Datos; se buscará definir los aspectos principales de esta disciplina y su aplicación en los medios de comunicación.

La intencionalidad general de esta investigación es definir qué es el Periodismo de Datos y generar una visibilidad del mismo para que empiece a ser utilizado con mayor asiduidad en los medios de comunicación tanto regionales como nacionales.

Índice

Índice	3
Introducción	5
Composición del trabajo	7
Capítulo I	8
¿QUÉ ES EL PERIODISMO DE DATOS?	8
Definición.....	8
Por qué es importante el periodismo de datos.....	12
Definición de “datos”	12
HISTORIA DEL PERIODISMO DE DATOS	14
Periodismo de Precisión.....	15
Periodismo Asistido por Computadoras.....	16
LA PIRAMIDE INVERTIDA DE PAUL BRADSHAW.....	17
PERFIL DEL PERIODISTA DE DATOS O <i>DATA JOURNALIST</i>	24
LA VISUALIZACION: LA HERRAMIENTA MAS FUERTE DEL PERIODISMO DE DATOS	25
EL PAPEL DEL BIG DATA.....	27
¿De dónde viene esa información?	28
El movimiento Open Data.....	31
El panorama en Argentina.....	33
Capítulo II	34
LOS DATOS COMO HERRAMIENTA PERIODISTICA	34
Cómo se utilizan y obtienen los datos.....	34
Cuatro productos del Periodismo de Datos.....	37
PRINCIPALES HERRAMIENTAS UTILIZADAS	38
Herramientas de recolección.....	39
Herramientas de filtro	41
Herramientas de visualización	42
Capítulo III	44
ENTIDADES DEL PERIODISMO DE DATOS	44
Knight Foundation	44
Hacks and Hackers.....	45
Wikileaks.....	46
PIONEROS DEL PERIODISMO DE DATOS	48
En el mundo: El caso The Guardian	48
En Argentina: El caso La Nación Data	50
Capítulo IV	52
ALGUNOS EJEMPLOS DE PERIODISMO DE DATOS	52
Datablog de The Guardian	52
La Nación Data	58

Hacks/Hackers capítulo Buenos Aires: Electoral Hack	63
Capítulo V	66
PERIODISMO DE DATOS EN ROSARIO.....	66
EL CASO DE PORVEN.....	66
Mapa de hechos.....	68
Mapa de cortes programados	71
EL CASO DE HACKS/HACKERS ROSARIO.....	79
Mapa de homicidios dolosos.....	81
Conclusión	90
Bibliografía	93
Anexo	95
ENTREVISTAS PERSONALES	95
Entrevista a Horacio Bella – Diario Porven.....	95
Entrevista a Ezequiel Clerici – Hacks/Hackers Rosario	105

Introducción

“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”
– Giannina Segnini

El reciente incremento de las nuevas tecnologías de la comunicación, la influencia de Internet en la producción cada vez más vertiginosa de datos, el surgimiento del movimiento Open Data, la implementación del trabajo conjunto de programadores, periodistas y visualizadores; así como la producción de software de código abierto, el proyecto de Gobierno Abierto y el desarrollo ágil de los dispositivos móviles; han transformado la práctica del periodismo, generando una nueva disciplina en auge: el Periodismo de Datos.

En este momento, casi la totalidad de la información se encuentra en formato digital, los datos abundan y exigen ser analizados. La cantidad de datos producidos en Internet es monstruosa, y aumenta cada vez más. Gracias a esto han surgido nuevas herramientas digitales para el tratamiento y el uso de datos producidos en la web, nuevos procesos de comunicación y nuevos usos de software para la producción de contenidos específicos.

El incremento de la cantidad de datos que dispone la sociedad en su conjunto, provocó la exigencia a los medios de comunicación de aumentar el rigor y la profesionalidad de sus publicaciones. Los datos ya no son entregados al público como algo dado, sino que comienzan a ser analizados, procesados y visualizados de tal manera que el usuario pueda hacerse con esos datos y llegar a generar sus propias conclusiones.

Ahora, estos grandes volúmenes de datos, impulsados por Internet, hacen que la disciplina denominada “Periodismo de Datos” comience a desarrollarse en mayor medida, irrumpa en las redacciones de los medios digitales y se consolide como una herramienta eficaz para lograr una mayor credibilidad en el contenido.

En este trabajo, por lo tanto, se intentará explicar qué es el Periodismo de Datos, aproximándonos a una definición estimativa del mismo, para poder determinar qué función cumple en el desarrollo de noticias e investigaciones basadas en datos.

La propuesta es indagar en los conceptos del Periodismo de Datos, analizando las principales herramientas que utiliza para recolectar, analizar y visualizar los datos.

Respecto de los objetivos secundarios y específicos de este trabajo, se detallarán las características de esta disciplina, se definirá el rol que cumple el Periodismo de Datos en la producción de noticias e investigaciones a partir de la utilización de herramientas digitales y software de código abierto, y se especificará de qué manera el Periodismo de Datos utiliza herramientas digitales y software específico para llevar a cabo su actividad.

Respecto de este último punto, se introducen ejemplos de proyectos de Periodismo de Datos realizados por las principales entidades y medios que trabajan con esta disciplina a nivel mundial, para ver de primera mano cómo se desarrolla y cómo son sus productos finales.

A fin de conocer en profundidad qué es el Periodismo de Datos y qué papel cumple en la producción de contenidos de comunicación, en este trabajo se hace hincapié en dos organizaciones de la ciudad de Rosario que utilizan a los datos como herramientas de información: el diario digital Porven y la organización Hacks/Hackers Rosario. Para esto se han realizado entrevistas a sus principales representantes y se han analizado sus respectivos trabajos de Periodismo de Datos, para poder responder a la hipótesis de este trabajo, la cual plantea que “el Periodismo de Datos cumple un papel importante en la producción de noticias a partir de la utilización de herramientas digitales y software”.

La elección de estos medios para analizar esta disciplina en auge está determinada por la influencia que tienen los mismos respecto del tema en la ciudad de Rosario. Al ser las dos únicas organizaciones que se desempeñan en el Periodismo de Datos en la ciudad, son de fundamental importancia para responder a las preguntas y objetivos planteados en este trabajo.

El valor de la presente publicación radica en contribuir a la comprensión de esta disciplina y a la visualización de la misma ante los medios de comunicación hegemónicos. En la ciudad de Rosario son pocos los medios que utilizan a los datos como herramientas periodísticas en el sentido en que lo hace el Periodismo de Datos, por eso es menester tener en cuenta esta disciplina para generar noticias marcadas por la rigurosidad periodística tanto ponderada y necesaria, y por las visualizaciones de datos que ofrecen un panorama nuevo y casi inexplorado dentro del periodismo.

Composición del trabajo

En el **Capítulo 1** de esta investigación se trata de hacer una aproximación a una definición de Periodismo de Datos y a sus principales conceptos, recorriendo su historia y los motivos por los cuales es necesaria esta disciplina. Aquí también se define el perfil del periodista de datos y los resultados finales de esta especialización. De la misma manera se contextualiza al Periodismo de Datos con la introducción de la definición del Big Data y el movimiento Open Data. Este capítulo tiene el fin de empezar a introducir conceptos de Periodismo de Datos para su mayor comprensión.

En el **Capítulo 2** se hace hincapié en la concepción de los datos como herramientas para producir noticias e investigaciones. El propósito de éste es detallar qué herramientas digitales se utilizan en el Periodismo de Datos y con qué objetivo.

En el **Capítulo 3** de esta investigación se detallan y numeran las principales entidades del Periodismo de Datos en el mundo y en Argentina, para poner en contexto el desarrollo de esta disciplina.

En el **Capítulo 4** se introducen ejemplos de Periodismo de Datos, analizados y visualizados para alcanzar una mayor comprensión de esta disciplina, con el objetivo de ejemplificar qué productos de Periodismo de Datos se pueden llevar a cabo.

En el **Capítulo 5** se desarrolla el eje principal de la investigación, donde se introduce a las dos entidades que trabajan con Periodismo de Datos en la ciudad de Rosario (el diario digital Porven y la organización Hacks/Hackers Rosario). También se analizan ejemplos de los productos de Periodismo de Datos realizados por estas entidades, cómo se han llevado a cabo y con qué herramientas, tomando como referencia los conceptos planteados en los capítulos anteriores.

Capítulo I

¿QUÉ ES EL PERIODISMO DE DATOS?

“El Periodismo de Datos es la práctica de encontrar noticias en los números y usar los números para narrar noticias” – Meredith Broussard.

Definición

No existe, en este momento, una definición exacta de Periodismo de Datos, ya que ésta es una disciplina emergente; podría decirse: una disciplina en pañales.

En el ámbito académico es una especialización periodística poco investigada; principalmente se ha desarrollado en el ámbito profesional, por lo que no existe una definición exacta del concepto de “Periodismo de Datos”. A pesar de esto, en este trabajo vamos a tratar de definir qué es el Periodismo de Datos y qué función cumple en la producción de proyectos periodísticos.

La consolidación de Internet como el principal medio para la obtención de información, ha proporcionado al periodismo innumerables herramientas para llevar a cabo su actividad. Se ha convertido en la musa y Dios de todos aquellos para los cuales la información es su fuente de trabajo.

Esa información está dada y ahora necesita ser interpretada.

Gracias a Internet, a las iniciativas de Gobierno Abierto y al Open Data, el incremento vertiginoso de la cantidad de datos en la web y a un creciente uso de tecnologías móviles: la información de la que disponen los periodistas es cada vez más abundante. Por eso, su sistematización, análisis e interpretación se ha vuelto crucial.

Podría decirse que el origen “oficial” del Periodismo de Datos y su definición fue dado en 1973 por Philip Meyer, uno de los principales representantes de esta disciplina. Meyer definió al Periodismo de Datos como “la aplicación de métodos científicos de investigación social y comportamental a la práctica del periodismo”¹, a lo que se le suman las

¹ Sabés Turmo, Fernando, y José Juan Verón Lassa. *Comunicación y la red. Nuevas formas de periodismo*. Zaragoza: Asociación de Periodistas de Aragón.

herramientas digitales y los grandes volúmenes de datos disponibles en la web, con los que trabaja esta disciplina.

El Periodismo de Datos es una especialización periodística que consiste en la búsqueda de noticias e investigaciones a través del análisis de grandes volúmenes de datos; es posible gracias al acceso a la información proporcionado por Internet. Consiste en la recolección, análisis y visualización de grandes cantidades de información, de manera fácil y sencilla de interpretar por el usuario. La tarea del periodista de datos es hacer comprensible esta información a la población mediante el análisis y visualización de los datos.

Para comenzar a definir un poco de qué se trata esta disciplina, es preciso destacar que el Periodismo de Datos tiene tres características principales:

- Trabaja con grandes volúmenes de datos, que muchas veces ni siquiera pueden alojarse en una hoja de cálculo de Excel.
- Su herramienta más importante es la visualización interactiva.
- Se incorpora al programador y diseñador al equipo periodístico para desarrollar tareas².

El trabajo en equipos multidisciplinarios es fundamental en esta rama periodística, es uno de sus ejes. Un periodista de datos puede ser tanto un profesional del periodismo como un desarrollador, programador o visualizador. Los equipos de trabajo del Periodismo de Datos se enriquecen con el trabajo en conjunto, participando todos sus miembros de las distintas etapas del proceso de investigación.

Es preciso aclarar que el Periodismo de Datos –a pesar de ser una disciplina relativamente reciente- tiene sus bases en el periodismo de investigación, que, según Sandra Crucianelli, “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”³.

El Periodismo de Datos trabaja con bases de datos que muchas veces son tan grandes que es difícil analizarlas según los métodos convencionales del periodismo de investigación, por eso, el aspecto innovador de esta disciplina en auge es el de poder

² Crucianelli, Sandra. *Herramientas Digitales para Periodistas*. 2013.

³ Perry, Felipe, y Miguel Paz. *Manual de Periodismo de Datos Iberoamericano*. 2013.

analizar, limpiar y procesar cantidades de datos monstruosas, mediante herramientas digitales que están a disposición del periodista en la web.

Es preciso aclarar que gran parte del mérito del Periodismo de Datos es encontrar relaciones entre las variables de esos datos que a simple vista pueden parecer inexistentes, pero que con la utilización de la cantidad completa del conjunto de datos (*datasets*) y de las herramientas disponibles, pueden producirse relaciones que conformen historias entre estos sucesos; sucesos que en un principio podrían parecer aislados, pero que en su conjunto conforman una base sólida de información. El mérito no está en encontrar los datos, sino en saber interpretarlos.

Sin embargo, esta capacidad de extraer y trabajar con grandes cantidades de información para después limpiarla y sistematizarla no es, aún, Periodismo de Datos. La sistematización de la información y el trabajo con datos ya es una práctica utilizada por el periodismo de investigación y por la investigación científica; conseguir los *datasets* (conjuntos de datos) o acceder a la información es fácil y cualquiera puede hacerlo. Pero para hacer Periodismo de Datos es preciso brindar un servicio de interés público, el cual es el fin último del periodismo. No basta con tener los datos: es preciso analizarlos y visualizarlos para encontrar asociaciones y relaciones que con anterioridad no se han visto, para hacerlas públicas de forma que cumplan con este propósito de brindar un servicio público.

Para aclarar este aspecto, es preciso mencionar que todo el periodismo tiene como base de datos la información, pero, como expresan Jesús Flores Vivar y Cecilia Salinas Aguilar en su trabajo “El periodismo de datos como especialización de las organizaciones de noticias en Internet”, al utilizar la palabra *datos* o *data* “implícitamente nos referimos a un tipo de información particular que puede ser procesada por sistemas informáticos y potentes herramientas de computación”⁴.

El Periodismo de Datos produce crónicas cortas con conclusiones contundentes, sustentadas mediante una visualización interactiva que muchas veces es el producto final y único de la investigación. Es por esto que el periodista de datos requiere de una capacitación constante, para poder manejar críticamente los datos, proporcionándoles un contexto y aplicándole procesos de análisis y verificación. A la hora de realizar este tipo de proyectos, es preciso y necesario el trabajo en conjunto de programadores y visualizadores.

⁴ Flores Vivar, Jesús, y Cecilia Salinas Aguilar. *El periodismo de datos como especialización de las organizaciones de noticias en Internet*. Universidad Complutense de Madrid, 2013.

Mediante las visualizaciones, el periodista es capaz de exponer una noticia o investigación de manera interactiva, dándole al lector la posibilidad de encontrar su propia historia en los datos. Se produce una re-significación de las infografías, las cuales pasan de ser elementos estáticos a productos maleables según la necesidad del usuario.

Una de las características principales del Periodismo de Datos es la publicación de las bases de datos en crudo –ya sea en las visualizaciones finales, en el artículo basado en datos o en las aplicaciones de noticias- para que el usuario tenga la posibilidad de navegar en los datos, re-significando su contenido y encontrando nuevas historias por su cuenta. Este aspecto del Periodismo de Datos es fundamental para responder a la necesidad cada vez más creciente de objetividad y credibilidad de la información.

En este sentido, el Periodismo de Datos le da un rigor periodístico a sus investigaciones que las noticias convencionales basadas en notas de prensa no pueden aportar. Esta disciplina responde a una necesidad creciente de confianza en la información, en un momento en que la credibilidad de los medios de comunicación está en tela de juicio.

Entonces: ¿qué es el Periodismo de Datos?

La definición que yo elijo dar es la siguiente: el Periodismo de Datos es una disciplina periodística, derivada del Periodismo de Investigación, que produce noticias e investigaciones mediante la utilización de grandes volúmenes de datos –representados en *datasets*-, mediante la utilización de herramientas digitales y software específicos proveídos por la Web. Estos datos, una vez extraídos son analizados, procesados, limpiados y sistematizados para luego ser volcados en un producto de Periodismo de Datos – preferentemente en una visualización interactiva-, para proporcionarle a la audiencia una historia contada con datos. Una característica fundamental de esta disciplina es que los datos en crudo son publicados junto con los productos finales del Periodismo de Datos, para que el usuario pueda explorar las bases de datos y encontrar sus propias historias.

Para terminar de definir qué es el Periodismo de Datos, es preciso mencionar lo que opina Paul Bradshaw⁵ al respecto:

“¿Qué hace que el Periodismo de Datos sea diferente al resto del periodismo? Tal vez son las nuevas posibilidades que se abren cuando se combina la tradicional ‘nariz para las noticias’ y la capacidad de contar una historia convincente, con la gran escala y gama de información digital ahora disponible”. (Bradshaw, Scraping for Journalists 2013)

⁵ Chaparro Domínguez, María Ángeles. *La evolución del periodismo de precisión: el blog de The Guardian sobre periodismo de datos*. Logroño: Universidad Internacional de La Rioja, 2013.

Por qué es importante el periodismo de datos

Una vez definido qué es el Periodismo de Datos, es preciso entender por qué esta disciplina cumple un papel tan importante en la producción de información periodística.

Anteriormente mencionábamos que la credibilidad en sus fuentes que brinda el Periodismo de Datos es superior y casi inigualable comparada con otras formas de hacer periodismo. Esto es importante, ya que nos encontramos en un momento de la historia digital donde el volumen de datos que se encuentran en la web es tan grande que muchas veces es imposible de entender.

Como mencionábamos anteriormente, el Periodismo de Datos es capaz de encontrar relaciones entre esos datos que a menudos son imperceptibles a simple vista y que sólo pueden ser comprendidas mediante su análisis con herramientas digitales.

Es justamente por esto que Mirko Lorenz asegura que “los periodistas deberían ver los datos como una oportunidad”⁶, ya que “usar datos transforma algo abstracto en algo que todos pueden entender y con lo que pueden relacionarse”. De esto hablábamos cuando mencionábamos que el periodismo, más específicamente el Periodismo de Datos, cumple un rol de servicio público.

La creciente importancia de esta disciplina radica en la posibilidad de dar contexto y claridad a la cantidad exorbitante de datos que circulan por la web, en descubrir qué es cierto y qué no, interpelando a la fuente misma de información.

Los periodistas que utilicen datos de la manera en la que lo hace esta disciplina, encontrarán una forma eficaz de apoyar sus artículos en datos y crear su propia posición frente a los hechos sin necesidad de un mediador. A su vez, le aportarán una credibilidad a sus investigaciones que de otras formas es muy difícil de asegurar.

Definición de “datos”

Un concepto clave del Periodismo de Datos es la fuente con la que trabaja, es decir: los datos. Pero, ¿qué son los datos?

Según la RAE un “dato” puede ser tres cosas:

1. “Información sobre algo concreto que permite su conocimiento exacto o sirve para deducir las consecuencias derivadas de un hecho. *A este problema le faltan datos numéricos*”.

⁶ «The Data Journalism Handbook.» 2011.

2. “Documento, testimonio, fundamento”.
3. “*Inform.* Información dispuesta a manera adecuada para su tratamiento por una computadora”.

Vamos a quedarnos con la última definición. Un dato es “información dispuesta a manera adecuada para su tratamiento por computadora”, es decir, procesable, cuantificable e interpretable.

El Periodismo de Datos se basa, como expresa su nombre, en datos. Éstos pueden ser la fuente o la herramienta para narrar historias. Es información, a menudo irrelevante, que individualmente pueden no significar nada pero que en su conjunto representan una fuente inagotable de información útil para los periodistas⁷. Sólo es preciso saber dónde encontrarlos y cómo interpretarlos.

Los datos se encuentran en todas partes, en los códigos de barra, en las antenas telefónicas, en las tarjetas de crédito, en los sensores de humedad, en las cajas registradoras, etc. Pero es Internet en donde se produce la mayor cantidad de datos, principalmente en Gmail, Google Docs, Facebook, Twitter, Instagram, Tumblr; todas las redes sociales o

“Los datos pueden ser la fuente del Periodismo de Datos, o pueden ser la herramienta con la que se narra la historia, o ambas cosas” (The Data Journalism Handbook 2011).

medios de interacción que usamos a diario y a los que creemos que no le estamos revelando más de nuestra vida de lo que les queremos revelar.

Hassel Fallas define a los *datos* como “una representación simbólica (numérica, alfabética, algorítmica, espacial, etc.) de un atributo o variable cuantitativa o cualitativa”⁸. El dato debe ser estructurado y comprensible para ser procesado por una computadora, claro está,

siendo un código binario.

Por ende, la definición de *datos* que yo elijo dar es la siguiente:

Los datos, en su conjunto, son información relevante para contar historias. Pueden ser usados como herramienta o como fuente para los periodistas, siempre y cuando estén en un

⁷ «The Data Journalism Handbook.» 2011.

⁸ Fallas, Hassel. «Tutorial: limpieza y análisis de datos.»

formato que pueda ser procesado por computadora, estructurado, cuantificable e interpretable. Un dato es una fuente inagotable de información periodística.

HISTORIA DEL PERIODISMO DE DATOS

La tarea de utilizar datos para mejorar noticias o informes y presentar la información de manera estructurada al público tiene una larga historia. Al contrario de lo que podríamos imaginar, el Periodismo de Datos no es reciente, o por lo menos no en su forma actual.

El periodismo fundado en bases de datos comenzó a implementarse hace más de tres décadas, en las redacciones (mayoritariamente estadounidenses) mediante micro-computadoras que se volvieron comunes en la década de 1980 (¡gracias Steve Jobs!).

Pero sin duda, lo que dio nacimiento al Periodismo de Datos fue la llegada de Internet. Nuestro querido y bendito Internet. Pero antes de ahondar en este aspecto, es preciso remontarse en el tiempo, al primer tercio del siglo XX. En este siglo, según explica Ana María Ávila, existía la preocupación latente de hacer del periodismo un oficio medible⁹; es por esto que se empezó a relacionar al trabajo del periodista con el del científico social. Uno de los principales pioneros en este asunto fue Walter Lippmann quien planteaba que debían examinarse las relaciones humanas con instrumentos y métodos científicos.

Ana María Ávila expresa que el siguiente paso de la historia del Periodismo de Datos se dio en los '30 “cuando se incorporó en la Universidad de Wisconsin un programa de doctorado especializado en periodismo dentro del área de Ciencias Políticas y Sociología”¹⁰, por lo que el periodismo comenzó a relacionarse más hondamente con estas disciplinas. Ávila aclara que “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”¹¹. ¡Ah!, acá ya surge el término de *base de datos*.

El trabajo de Meyer está marcado por su investigación, en 1967, de las marchas y disturbios causados en Detroit, con la cual logró demostrar que, al contrario de lo que se pensaba, los manifestantes no sólo eran personas con baja escolaridad o de sectores

⁹ «The Data Journalism Handbook.» 2011.

¹⁰«The Data Journalism Handbook.» 2011.

¹¹ «The Data Journalism Handbook.» 2011.

marginales, sino que también protestaban personas que tenían estudios universitarios y que eran de clase media. Philip Meyer ganó un premio Pulitzer por este trabajo, y, tal como menciona Ávila en el “Manual de Periodismo de Datos Iberoamericano”, así surgió el periodismo de precisión, el precursor de lo que hoy se conoce como Periodismo de Datos¹².

Luego del legado de Philip Meyer, medios como The Washington Post o el Philadelphia Inquirer, según menciona Ávila, comenzaron a realizar investigaciones de la misma índole. “A finales de los años 70 y principio de los 80, los periódicos Philadelphia Inquirer, The Dallas Morning News, The New York Times, Los Ángeles Times, entre otros, ya practicaban el periodismo de precisión y ganaban premios con ello”¹³.

En el mundo de habla hispana, aclara Ávila, las noticias relacionadas con el periodismo de precisión comenzaron a realizarse a partir del 1992. ¡Bastante tarde! El diario El País de España fue el pionero en este asunto con “un texto sobre el tema bajo la autoría de Pedro Gómez y José Luis Dader”¹⁴.

Y de los países latinoamericanos ni hablemos, porque las posibilidades de aperturas de datos y libertad de expresión eran y son muy reducidas, un tema que vamos a analizar más adelante.

Para entender un poco más el contexto en el cual surgió el Periodismo de Datos es preciso ahondar en dos grandes corrientes: el Periodismo de Precisión y el Periodismo Asistido por Computadoras. Antecedentes de la disciplina que estamos analizando.

Periodismo de Precisión

Anteriormente mencionamos a Philip Meyer, quien acuñó este término propiamente dicho. Según Meyer, “a comienzos de la década del ’70 el término *periodismo de precisión* fue acuñado para describir este tipo de recolección de noticias: ‘la aplicación de métodos de investigación de las ciencias sociales y de la conducta a la práctica del periodismo’”¹⁵.

Meyer plantea que el Periodismo de Precisión nació en respuesta al famoso “nuevo periodismo”, y sugiere que lo que necesita el periodismo son técnicas científicas de recolección y análisis de datos, en lugar de técnicas literarias (caracterizadas por el “nuevo periodismo”).

¿Entonces qué es el Periodismo de Precisión?

¹² «The Data Journalism Handbook.» 2011.

¹³ «The Data Journalism Handbook.» 2011.

¹⁴ «The Data Journalism Handbook.» 2011.

¹⁵ Meyer, Philip. *The New Precision Journalism*. 1991.

El Periodismo de Precisión es la forma de combinar trabajo periodístico con trabajo científico, de recabar y analizar información mediante el método científico de análisis y recolección de datos.

Periodismo Asistido por Computadoras

Primero fue el Periodismo de Precisión, luego el Periodismo Asistido por Computadoras.

El Periodismo Asistido por Computadoras, conocido por las siglas CAR, fue “el primer abordaje organizado y sistemático del uso de computadoras para recoger y analizar datos”¹⁶, para mejorar la producción de noticias.

Una de la primeras veces en las que se utilizó el CAR fue en 1975 por la cadena CBS para predecir los resultados de la elección electoral de ese mismo año (¿Les suena a Watergate?)

Las prácticas del CAR se remontan a la segunda mitad del siglo XX; esta práctica, sin embargo, “ha cambiado con el tiempo junto con el cambio en las herramientas y el medio del mundo digital. Comenzó en los tiempos de los *mainframes* a fines de la década de 1960 y avanzó a las PC (que incrementaron la velocidad y la flexibilidad del análisis y la presentación) y luego a la Red, lo que aceleró la capacidad de reunir, analizar y presentar datos”¹⁷.

Hay quienes mantienen que existe una diferencia entre CAR y el Periodismo de Datos, alegando que CAR es una “técnica para recoger y analizar datos como una manera de fortalecer el periodismo”¹⁸, mientras que lo que conocemos como *Periodismo de Datos* se basa en la manera en que los datos “se ubican en el conjunto del flujo de trabajo periodístico”. Es por esto que el Periodismo de Datos hace hincapié en los datos como una fuente en lugar de utilizarlos meramente como medio para encontrar o dar más fuerza a las historias. En esto se diferencia el Periodismo de Datos del CAR o del Periodismo de precisión. En el *Periodismo de Datos* los datos pueden ser una fuente para la investigación o la historia misma.

¹⁶ «The Data Journalism Handbook.» 2011.

¹⁷ Howard, Alexander Benjamin. *El arte y la ciencia del Periodismo de Datos*. Columbia Journalism School.

¹⁸ «The Data Journalism Handbook.» 2011.

LA PIRAMIDE INVERTIDA DE PAUL BRADSHAW

Ya hemos definido qué es el Periodismo de Datos, hemos ahondado un poco en su historia y hemos justificado su necesidad. Ahora es preciso comenzar a introducir conceptos más específicos de esta disciplina, como son las partes del proceso que llevan a convertir a los datos en crudo en noticias de interés público.

Para esto vamos a comenzar explicando el concepto de “pirámide invertida” que formuló Paul Bradshaw, para contextualizar y explicar cómo trabaja el Periodismo de Datos.

Paul Bradshaw es un periodista británico especializado en periodismo digital, especialmente en blogs y Periodismo de Datos. Es profesor en Birmingham City University y en la City University of London. Principalmente (a lo que concierne nuestro tema) es el creador del Online Journalism Blog, un blog que abarca temas como la producción de análisis, comentarios y links que conciernen al periodismo digital, el Periodismo de Datos, los blogs, los vlogs (que no es lo mismo), el fotoblogging, los podcast, el storytelling, etcétera, etcétera. En definitiva: abarca todo el mundo virtual destinado al periodismo. Bradshaw también es colaborador del Datablog de The Guardian, tema que vamos a comentar más adelante.

Este periodista planteó, en su blog Online Journalism Blog, una forma de ver al Periodismo de Datos como una “pirámide invertida”, que es la siguiente:

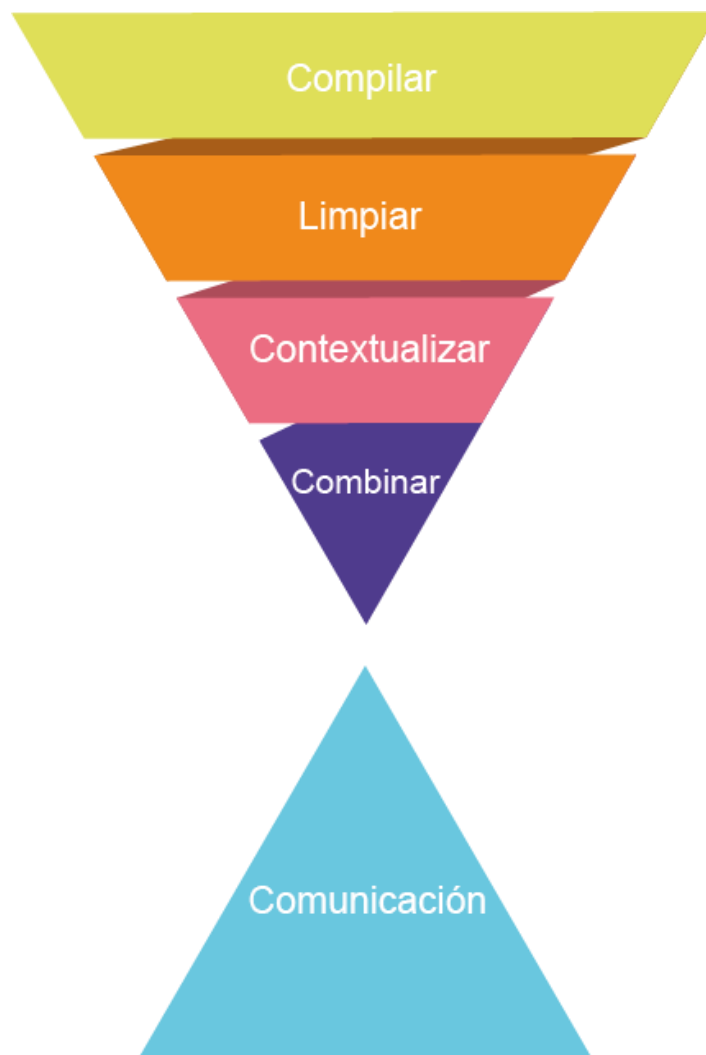


Ilustración 1

La pirámide de Paul Bradshaw muestra cómo, desde una cantidad enorme de datos, se va depurando la información hasta convertirse en un producto de Periodismo de Datos que sea inteligible para la audiencia, de la manera más sencilla y práctica posible. Es decir, planteó las bases de lo que podría ser una aproximación a una teoría de Periodismo de Datos¹⁹.

Los elementos que componen la pirámide son los siguientes:

- 1. Compilar**

Este autor plantea que el Periodismo de Datos puede comenzar de dos formas: o se tiene una pregunta que necesita *data* para ser contestada o se tiene un *dataset* (conjunto de datos) que necesita ser interrogado.

¹⁹ Bradshaw, Paul. *onlinejournalismblog.com*. <https://onlinejournalismblog.com/2011/07/08/the-inverted-pyramid-of-data-journalism-in-spanish/> (último acceso: Abril de 2017).

Con “interrogado” se refiere al proceso de encontrar la noticia en los datos, en hacerle preguntas a esos datos.

Bradshaw asegura que la compilación de la información es lo que lo define como un acto de Periodismo de Datos, el cual puede tener varias formas:

- Los datos pueden ser provistos directamente por una organización.
- Pueden ser encontrados usando técnicas de búsqueda avanzada para buscar estos datos en los sitios web de los gobiernos.
- Mediante *scraping* de bases de datos.
- Mediante la conversión de documentos a un formato que pueda ser analizado por las herramientas del Periodismo de Datos.
- Recolectando información de APIs.
- El propio periodista puede recolectar datos a través de la observación, mediante encuestas, realizando formularios online o por *crowdsourcing*.

2. Limpiar

Esta etapa de la pirámide invertida hace hincapié en la necesidad de confiar plenamente en la data que se está analizando, para ello es necesario limpiarla, lo cual involucra el hecho de reducir al mínimo (a cero si es posible) el error humano y convertir la data en un formato que sea compatible con el resto del material que se esté utilizando. En caso de que la data a utilizar se encuentre en un formato que no pueda ser procesado fácilmente por las herramientas del Periodismo de Datos, es necesario convertirla a un formato inteligible para éstas.

Hay formas simples de limpiar los datos en plataformas como Excel o Google Docs (dos herramientas fundamentales del análisis de datos).

3. Contexto

Paul Bradshaw plantea, en esta etapa del proceso, que los datos -como cualquier fuente- no pueden ser información siempre confiable. Los datos “vienen con sus propias historias, prejuicios y objetivos”²⁰. Por eso es necesario interrogarlos. Éstas son las preguntas que el autor recomienda hacer a los datos:

- ¿Quién recolectó los datos?
- ¿Cuándo y por qué motivos fueron recolectados?

²⁰ Bradshaw, Paul. *onlinejournalismblog.com*. <https://onlinejournalismblog.com/2011/07/08/the-inverted-pyramid-of-data-journalism-in-spanish/> (último acceso: Abril de 2017).

- ¿Cuál fue la metodología utilizada para la recolección?

También es necesario entender la jerga: los códigos que representan categorías, clasificaciones o ubicaciones, así también la terminología especializada.

Estas preguntas llevan a la búsqueda de más información, al incremento de la base de datos que se va a utilizar. A su vez, tener una pregunta clara al comienzo de todo el proceso ayuda a asegurar que no se pierda el foco o el ángulo interesante de la investigación.

4. **Combinar**

Es posible encontrar una buena historia en un solo conjunto de datos, pero a veces estos datos no son suficientes y es necesario combinar dos conjuntos de datos, o más. Las noticias con múltiples fuentes tienen mayor valor informativo que las compuestas por una sola fuente.

Paul Bradshaw dice que la combinación más utilizada es el “mashup de mapas”: se toma un dataset y se lo combina con datos de mapas. Así se logra proveer una visualización instantánea de esos datos, de cómo algo está distribuido en el espacio. Una de las herramientas más utilizadas para este propósito es Google Maps, tan utilizada, dice Bradshaw, que se ha convertido en un cliché.

En esta parte del proceso de la pirámide invertida hay un sinnúmero de opciones para combinar datasets con otras herramientas, las cuales vamos a ver más adelante.

5. **Comunicar**

Esta es la última etapa del proceso, y la más importante. Es la etapa que más caracteriza al Periodismo de Datos: la visualización.

Una vez compilados, limpiados, contextualizados y combinados los datos, es preciso visualizar los resultados. Esto puede hacerse mediante mapas, cuadros interactivos, infografías, animaciones, o convertir ese dataset en aplicaciones de noticias.

Paul Bradshaw explica el proceso de comunicación con otra pirámide:



Ilustración 2

En la pirámide de **comunicación** Paul Bradshaw presenta cinco formas diferentes de comunicar en Periodismo de Datos²¹.

1. **Visualización**

La visualización es la forma más rápida y eficiente (predilecta) de comunicar los resultados del Periodismo de Datos. Existen herramientas gratuitas para llevar a cabo una visualización, como Google Docs, que son fáciles de utilizar, o herramientas más poderosas como Many Eyes donde el usuario “pega” la data cruda y selecciona un grupo de opciones de visualización.

Las visualizaciones deben llevarse a cabo de manera efectiva para que sean de fácil lectura para el usuario. Paul Bradshaw recomienda que para generar infografías hay que enfocarse en tres puntos: 1) no debe tener más de cuatro puntos de datos, 2) evitar las visualizaciones 3D, 3) asegurarse de que el gráfico sea autosuficiente.

²¹ Bradshaw, Paul. *onlinejournalismblog.com*. <https://onlinejournalismblog.com/2011/07/08/the-inverted-pyramid-of-data-journalism-in-spanish/> (último acceso: Abril de 2017).

Según este autor, la visualización tiene la gran ventaja de hacer que la comunicación sea efectiva. Provee un método de distribución de su contenido que no puede ser igualado por otros tipos de comunicación.

Bradshaw también recomienda asegurarse que la infografía contenga un enlace a la fuente y que haya información adicional en la fuente cuando la audiencia llegue a ella. Como mencionábamos anteriormente: publicar el conjunto de datos en bruto junto con la visualización.

2. Narración

Es necesario interpretar los datos y las visualizaciones realizadas en base a ellos, por eso la narración es un proceso necesario y fundamental. Una visualización siempre debe ser acompañada de un texto que la contextualice.

Las noticias generadas por el Periodismo de Datos suelen centrarse mayoritariamente en las visualizaciones, pero es necesario introducir texto y narración a este producto para que pueda ser mejor interpretado. Paul Bradshaw dice que, como con las visualizaciones, menos es más. Pero asegura que también se necesita pensar en el significado y los objetivos al comunicar los datos.

“La cifras abstractas pueden ser impresionantes, pero sin sentido e inútiles”²² recalca el autor. Por eso es preciso reducir los porcentajes, cifras o números que suelen aparecer en este tipo de noticias con datos, a cantidades manejables.

Es recomendable, también, utilizar las infografías para enfocarse en las cuestiones principales y enlazar los datos en crudo para cuestiones más específicas.

3. Comunicación social

Bradshaw asegura que la comunicación es un acto social, y que el éxito de las infografías a través de medios sociales es un testimonio de ello. La información también es social.

El Periodismo de Datos, según este autor, puede aprovechar la data social (obtenida mediante formatos como Facebook, Twitter, etc), y, al mismo tiempo, se pueden comunicar los resultados del Periodismo de Datos utilizando elementos como concursos, competiciones, campañas y colaboraciones.

Paul Bradshaw asegura que nos encontramos recién en el comienzo de este aspecto del periodismo online, y de la disciplina misma.

²² Bradshaw, Paul. *onlinejournalismblog.com*. <https://onlinejournalismblog.com/2011/07/08/the-inverted-pyramid-of-data-journalism-in-spanish/> (último acceso: Abril de 2017).

4. **Humanizar**

Gracias a la implementación de la tecnología y a la utilización de gráficos para contar historias, se ha dejado de lado la humanización del contenido. Paul Bradshaw señala que, en los programas de noticias, a menudo se utilizan casos de estudio para tratar el problema de presentar historias basadas en datos o números en medios como la televisión o la radio. Se humanizan los números.

A pesar de la enorme ventaja que representa generar noticias a partir de datos, (recolectados, depurados y publicados por computadora) es necesario, según este autor, salir a la calle y grabar entrevistas relacionadas a esos datos, para propulsar de manera más efectiva (y humanizada) la noticia.

5. **Personalizar**

Una de las principales características del periodismo online es la interactividad. En el Periodismo de Datos, dice Bradshaw, eso significa que los usuarios pueden, potencialmente, controlar qué información es presentada a ellos en varias entradas.

Una forma de interacción entre el usuario y la información puede ser los sitios interactivos, donde el usuario puede tomar sus propias decisiones respecto, por ejemplo, de cómo recortarían el déficit.

Otra forma muy común es la personalización geográfica, donde el usuario puede introducir su código postal, por ejemplo, para descubrir cómo un tema en particular está afectando o influenciando en su lugar de residencia.

Según este autor, las posibilidades de personalización se expanden, mientras más data personal esté en manos de terceros.

Bradshaw dice que “las noticias personalizadas pueden, de muchas maneras, ser utilizadas como una expresión de nuestra identidad: acá es donde vivo, de esta manera me afecta, en esto estoy interesado”²³.

Se tiene que mostrar al público su cercanía con los datos, cómo le afectan personalmente.

Acá es cuando el Periodismo de Datos cumple su función de servicio público.

²³ Bradshaw, Paul. *onlinejournalismblog.com*. <https://onlinejournalismblog.com/2011/07/08/the-inverted-pyramid-of-data-journalism-in-spanish/> (último acceso: Abril de 2017).

PERFIL DEL PERIODISTA DE DATOS O *DATA JOURNALIST*

El Periodismo de Datos, como mencionamos anteriormente, es una especialización emergente, por lo cual no se puede definir un perfil periodístico exacto, pero sí se pueden hacer aproximaciones al mismo.

Al Periodismo de Datos lo pueden hacer todos: un periodista, un programador, un diseñador, un curioso que empezó a probar con las herramientas de procesamiento de datos. Cualquiera. Pero hay ciertas características que son necesarias a la hora de manipular noticias con bases de datos, y a la hora de hacer Periodismo de Datos.

En el texto “Aproximación teórica al perfil profesional del ‘periodista de datos’”, Eva María Ferreras Rodríguez menciona la postura de Flores y Cebrián respecto del perfil del periodista de datos; estos autores creen que el periodista de datos es “un profesional que puede estar formado en documentación, periodismo o cualquier otra disciplina, que posee un aditivo de conocimiento necesario para la gestión de datos y clasificación de información, así como para la búsqueda, procesamiento, análisis de datos, etc”²⁴. Este periodista deberá ser capaz de manejar grandes volúmenes de información, poder clasificarlos y después volcar los datos en las herramientas tecnológicas disponibles.

Respecto de la formación que necesita poseer un periodista de datos, Flores y Cebrián consideran necesario que posea conocimientos básicos en campos diversos como la estadística, diseño, programación, sociología y psicología, además de sus conocimientos periodísticos.

Paul Bradshaw, un autor que ya mencionamos, señala que “los profesionales del Periodismo de Datos necesitan una serie de habilidades”²⁵, y destaca que lo más importante, sin embargo, es la capacidad de encontrar información y hacerla inteligible para el público.

El Periodismo de Datos requiere de una capacitación constante. Es esencial que los periodistas especializados en datos se capaciten “en el desarrollo de un manejo crítico de los datos, colocándolos en contexto y aplicando procesos de curaduría y verificación”²⁶.

²⁴ Ferreras Rodríguez, E. M. «Aproximación teórica al perfil profesional del ‘Periodista de datos’.» 2013.

²⁵ Bradshaw, Paul. «Scraping for Journalists.» Leanpub, 2013.

²⁶ Perry, Felipe, y Miguel Paz. *Manual de Periodismo de Datos Iberoamericano*. 2013.

Respecto de la capacitación constante, Sandra Crucianelli coincide con este aspecto y agrega que el periodista también tiene que tener conocimientos de materias que, generalmente, están alejadas de las preferencias del periodista convencional, tales como la matemática, la programación y la estadística. El Periodismo de Datos es una especialización que a simple vista parece relacionarse más con áreas como la matemática o el análisis de sistemas, hecho que puede desanimar a un periodista neófito en el tema; pero Crucianelli señala que la capacidad de reunir conocimientos variados para realizar esta actividad es muy productiva. En cuanto a las habilidades que debe poseer el periodista de datos, destaca: “el conocimiento del funcionamiento de la administración pública, dominio del inglés, habilidades de búsqueda avanzada en la web, conocer los principios de programas de análisis de datos (*Excel* o similar), conocimientos de matemática básica y de estadística, así como de visualización”²⁷.

Por otro lado, en “La navaja suiza del reportero” se hace hincapié en que los periodistas necesitan aprender a programar. Si bien los periodistas especializados en datos suelen trabajar en colaboración con un programador, en este artículo se menciona la necesidad de desarrollar “mentalidades de datos masivos”, para que el periodista entienda que los datos contienen historias que deben ser descubiertas²⁸.

En definitiva, un periodista de datos debe contar con una capacitación constante en diversos aspectos, ya sea tecnológicos, investigativos o de diseño. Cualquiera que tenga la habilidad e insistencia para encontrar una historia en los datos está haciendo Periodismo de Datos.

Hay que recalcar que para desempeñarse en esta disciplina es preciso –de forma obligatoria- trabajar en conjunto. La ayuda del programador o diseñador en esta tarea es fundamental, y no solo enriquece la investigación sino también al periodismo mismo.

LA VISUALIZACION: LA HERRAMIENTA MAS FUERTE DEL PERIODISMO DE DATOS

Una parte fundamental de esta disciplina basada en datos es la visualización. Habíamos mencionado que las investigaciones de Periodismo de Datos pueden devenir en distintos productos, y que el más importante de ellos es la visualización interactiva. Es la manera

²⁷ Ferreras Rodríguez, E. M. «Aproximación teórica al perfil profesional del ‘Periodista de datos’.» 2013.

²⁸ Hidalgo, David, y Fabiola Torres. «La navaja suiza del reportero.» Lima, Febrero de 2016.

más directa, simple y efectiva de reproducir una noticia realizada con un gran volumen de datos.

Una visualización de datos que esté bien diseñada le puede ofrecer al público una interpretación profunda, fácil de entender y acotada del hecho.

Como habíamos dicho con anterioridad, el Periodismo de Datos trabaja con un volumen muy grande de información, que, sin una visualización adecuada, sería imposible de comprender o ser notada a simple vista.

Sandra Crucianelli asegura que “la visualización de datos explica de manera comprensible las relaciones entre gran cantidad de información que se genera por distintas vías”²⁹. De esta forma es posible relacionar variables de manera que el usuario pueda ver estas variables representadas de forma explícita y comprensible.

Los gráficos le permiten al lector decidir qué desea ver y cómo, qué variables combinar y qué resultados obtener, le muestran los datos de una manera más fácil y efectiva.

Lev Manovich define a la visualización como “una representación gráfica que expone datos concretos a través de su representación visual”³⁰. Este autor, en su libro “What is visualization” expresa que el objetivo de las visualizaciones de la información - especialmente las del Periodismo de Datos- es el de “descubrir la estructura de un set de datos (generalmente muy amplio)”³¹. Es decir, con una visualización de la información bien diseñada, se logra descubrir una estructura en los datos que no es conocida a priori, pero que se revela de manera exitosa una vez visualizada.

Las visualizaciones del Periodismo de Datos son diferentes a las visualizaciones “primitivas” como las llama Lev Manovich. Aquellas como los gráficos de torta, gráficos de barra, gráficos de línea, etc. Éstos son los gráficos comunes, utilizados en las investigaciones científicas, periodísticas, etc. que ya no sirven para visualizar grandes volúmenes de datos.

Una de las formas de visualización que utiliza el Periodismo de Datos son herramientas como Word Cloud³² o Word Tree³³, donde la visualización es directa, sin reducción, en la que los datos se reorganizan dentro de la representación visual, preservando la forma

²⁹ Perry, Felipe, y Miguel Paz. *Manual de Periodismo de Datos Iberoamericano*. 2013.

³⁰ Manovich, Lev. *What is Visualization*. 2010.

³¹ Manovich, Lev. *What is Visualization*. 2010.

³² *Word Cloud*. <http://www.wordclouds.com/> (último acceso: Mayo de 2017).

³³ *Word Tree*. <https://www.jasondavies.com/wordtree/> (último acceso: Mayo de 2017).

original. Estas dos herramientas son una manera de volcar toda la base de datos y obtener resultados visuales inmediatos, que logran vislumbrar las relaciones entre esos datos.

Word Cloud es una herramienta online que genera una visualización interactiva, una “nube de palabras” justamente (también pueden ser etiquetas), que sirve para visualizar las palabras que más veces aparecen mencionadas en un texto. Esta nube se puede personalizar en su forma, tema, colores y tipografías. También se pueden filtrar las palabras, borrar los signos o caracteres que aparecen por defecto o que no son necesarios; se puede editar el tamaño de la nube y la separación entre las palabras.

De semejante manera funciona **Word Tree**, una herramienta de Google que permite introducir un texto (libro, artículo, noticia) y encontrar relaciones entre sus frases y palabras, y funciona de manera online. El árbol de palabras representa múltiples secuencias paralelas de palabras, y se puede usar de diferentes maneras: para saber qué palabras aparecen más seguido, para mostrar una jerarquía de términos, para saber qué frase precede a una palabra o viceversa. En esta herramienta se puede introducir una palabra clave que se quiera visualizar en el texto y de esta manera encontrar relaciones entre esta palabra y otras.

Estas dos herramientas responden al concepto que plantea Lev Manovich: la visualización sin reducción. Esta es una clase de visualización con la que trabaja el Periodismo de Datos y que “se refiere a la preservación de un conjunto mucho más rico de propiedades de los objetos de datos durante el proceso de creación de visualizaciones directamente de ellos”³⁴. Es decir: se conserva la base de datos entera, lo que supone una visualización directa de los mismos.

De esta manera, Lev Manovich asegura que “es posible revelar patrones manteniendo el 100% de la información”³⁵.

EL PAPEL DEL BIG DATA

¿Qué es el Big Data? Esta es una pregunta muy difícil de contestar, ya que no existe una definición exacta para este término. El Big Data vendría a ser toda aquella información que no puede ser procesada o analizada utilizando métodos o herramientas meramente tradicionales. Esta enorme cantidad de datos no sólo se define por su tamaño (representado por el *zettabyte*) sino también por su gran variedad de contenido. IBM afirma que “existe

³⁴ Manovich, Lev. *What is Visualization*. 2010.

³⁵ Manovich, Lev. *What is Visualization*. 2010.

una gran variedad de datos que pueden ser representados de diversas maneras en todo el mundo, por ejemplo de dispositivos móviles, audio, video, sistemas GPS, incontables sensores digitales en equipos industriales, automóviles, medidores eléctricos, veletas, anemómetros, etc.”³⁶ Esto es lo que mencionábamos anteriormente en nuestra definición de *datos*. Los datos están en todas partes, y ese mar de datos que no tiene fin y sigue creciendo cada día un poco más es denominado: *Big Data*.

Existen tres tipos de *Big Data*:

“Los big data –los datos masivos- se refieren a cosas que se pueden hacer a gran escala [...] para extraer nuevas percepciones o crear nuevas formas de valor, de tal manera que se transforman los mercados, las organizaciones, las relaciones entre ciudadanos y los gobiernos” (Hidalgo, David ; Torres, Fabiola 2016).

- *Datos estructurados*: son datos bien definidos en su formato, tales como fechas, números o cadenas de caracteres. Son usualmente almacenados en tablas como, por ejemplo, las hojas de cálculo.
- *Datos no estructurados*: son datos que están en el formato en el cual fueron recolectados, carecen de un formato específico y no se pueden almacenar en una tabla. Son, por ejemplo, el formato PDF, los documentos multimedia o documentos de texto.
- *Datos semiestructurados*: son datos que no tienen un campo determinado. Poseen sus propios metadatos semiestructurados. Son, por ejemplo, el HTML, XML o JSON.³⁷

¿De dónde viene esa información?

IBM afirma que los seres humanos almacenamos y creamos información constantemente, ya sea desde nuestros dispositivos tecnológicos personales o desde nuestras industrias y sistemas comunicacionales. De ahí y de muchos lugares más proviene la información que conforma el Big Data.

Esta empresa asegura que la información no solo proviene de los seres humanos. Dice que “existe también la comunicación denominada ‘máquina a máquina’ (M2M machine-to-machine) cuyo valor en la creación de grandes cantidades de datos también es muy

³⁶ IBM. <https://www.ibm.com/developerworks/ssa/local/im/que-es-big-data/> (último acceso: Mayo de 2017).

³⁷ Wikipedia. https://es.wikipedia.org/wiki/Big_data (último acceso: Mayo de 2017).

importante”³⁸. Un ejemplo que puede explicar este tipo de comunicación “del futuro” son los sensores digitales que están instalados en contenedores y que determinan la ruta generada durante una entrega de algún paquete; esa información es enviada a las compañías de transporte, a los sensores en medidores eléctricos y a las compañías del sector energético. Otro ejemplo podría ser la conexión e intercambio de datos entre celulares y computadoras, algo que ocurre diariamente.

Para entender aún más la cantidad astronómica de información que compone al *Big Data* hay que entender a qué cantidad de información equivale un zettabyte, la medida con la cual se caracteriza al Big Data, lo que se explica con este gráfico:



Ilustración 3 (Fuente: “La navaja suiza del reportero”)

Ahora sabemos que 1 *zettabyte* es una cantidad enorme de información.

Alicia Tapia afirma que la andadura de la era digital inició en el año 2002, “donde la capacidad de almacenamiento digital superó a la analógica: en 2007 casi el 94% de toda la

³⁸ IBM. <https://www.ibm.com/developerworks/ssa/local/im/que-es-big-data/> (último acceso: Mayo de 2017).

información estaba en formato digital”³⁹. A lo que hace referencia esta autora es al avance vertiginoso e imparable de la capacidad digital para almacenar información que de otra forma no podría ser almacenada sin ocupar cuantiosas cantidades de espacio físico.

Imagínense 1 *zettabyte* de información representado en papel... creo que la torre llegaría hasta la Luna. Eso es lo que expresa IBM cuando dice “si todos los bits y bytes de datos del último año fueran guardados en CD’s, se generaría una gran torre desde la Tierra hasta la Luna y de regreso”⁴⁰. Es muchísimo, eso está claro.

Alicia Tapia también menciona en su texto que toda, pero toda, la información digital se duplica cada dos años. Si esto es así, entonces razona que en 2010, solamente, la información alcanzaba y superaba la barrera del *zettabyte*. Imagínense cuántas torres a la Luna existirán para 2020. En la siguiente imagen se pueden ver distintos ejemplos de la cantidad de datos que representan las medidas que mencionamos anteriormente.

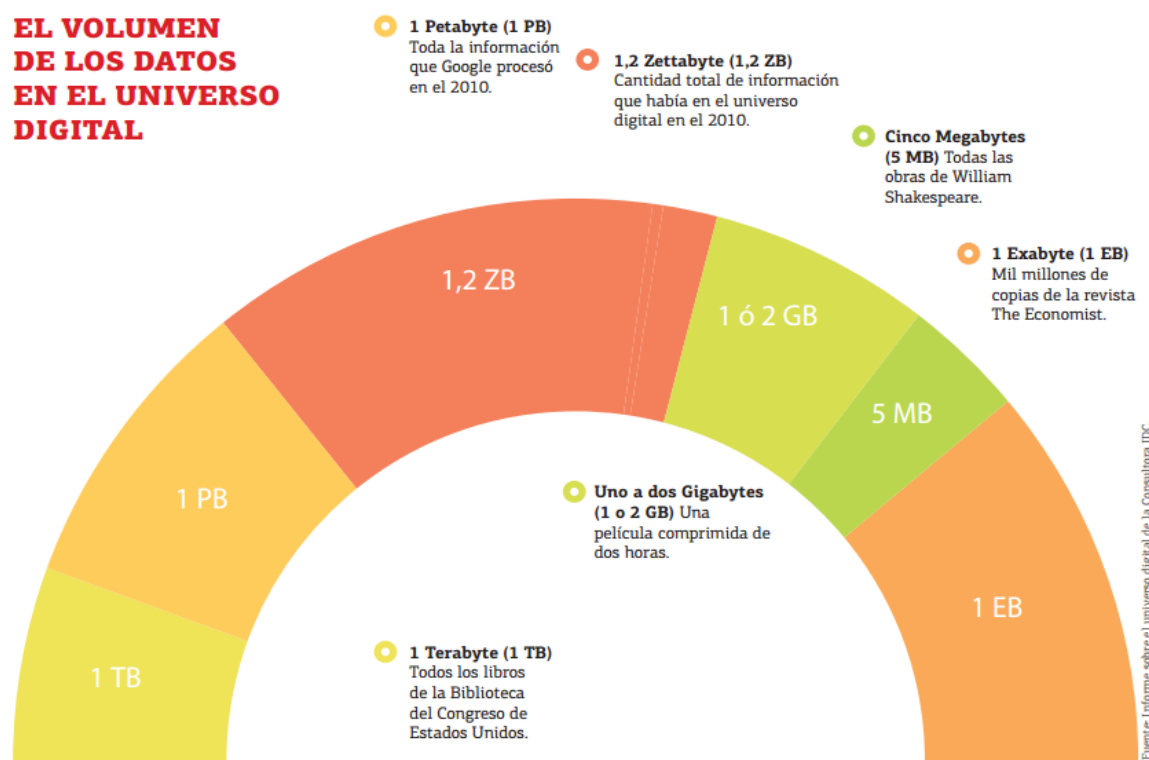


Ilustración 4 (Fuente: “La navaja suiza del reportero”)

Los datos han comenzado a acumularse a tal punto que se están convirtiendo en algo nuevo y especial, en un fenómeno de escalas enormes que crece a un ritmo vertiginoso,

³⁹ Tapia, Alicia. *Big Data y periodismo de datos*. Madrid: Universidad Internacional de La Rioja.

⁴⁰ IBM. <https://www.ibm.com/developerworks/ssa/local/im/que-es-big-data/> (último acceso: Mayo de 2017).

denominado: *Big Data*. El término “Big Data” o “Datos masivos” se acuñó en la década del 2000 gracias a la explosión de los datos en ciencias como la astronomía y la genética, acostumbradas a la utilización de grandes cantidades de datos para desarrollar sus investigaciones.⁴¹

No existe una definición rigurosa para este término, pero, los Big Data, los Datos Masivos, se refieren a cosas que se pueden hacer a gran escala, pero no a una escala inferior.⁴² Al utilizar grandes volúmenes de datos y emplearlos en su totalidad –el cual es el objetivo y fundamento del Periodismo de Datos- se pueden apreciar detalles o aspectos de los datos que no podrían ser vistos en muestras más pequeñas. Para entender estos datos, es preciso encontrar pautas y patrones en los mismos, que puedan guiarnos en las investigaciones y búsqueda de noticias.

“Jeff Jones, el experto en datos masivos de IBM, sostiene que hay que dejar que los datos ‘le hablen a uno’⁴³, es decir, que nos muestren el panorama completo de lo que se está investigando para que surjan las historias por sí solas.

Viktor Mayer-Schönberger y Kenneth Cukier, en su libro “Big Data. La revolución de los datos masivos”, aseguran que “el empleo de la totalidad de datos hace posible advertir conexiones y detalles que de otro modo quedan oscurecidos en la vastedad de la información. Disponer del conjunto de datos completo (o prácticamente completo) ofrece mucha más libertad para explorar, para estudiar los datos desde diferentes perspectivas, o para examinar más de cerca determinados aspectos”.⁴⁴

La consolidación del Big Data como la fuente más importante de información en la web, provocó la necesidad de transparencia a la hora de buscar y utilizar esos datos. Así fue cómo surgió el movimiento Open Data.

El movimiento Open Data

Con la implementación de la tecnología en las salas de redacción se han abierto nuevas fronteras para el periodismo, principalmente gracias a la cantidad enorme de datos que circulan en Internet, y que crece día a día. Gracias al aumento generalizado de la

⁴¹ Mayer-Schönberger, Viktor ; Cukier, Kenneth. «Big Data. La revolución de los datos masivos.» Madrid: Turner Publicaciones S.L, 2013.

⁴² Mayer-Schönberger, Viktor ; Cukier, Kenneth. «Big Data. La revolución de los datos masivos.» Madrid: Turner Publicaciones S.L, 2013.

⁴³ Mayer-Schönberger, Viktor ; Cukier, Kenneth. «Big Data. La revolución de los datos masivos.» Madrid: Turner Publicaciones S.L, 2013.

⁴⁴ Mayer-Schönberger, Viktor ; Cukier, Kenneth. «Big Data. La revolución de los datos masivos.» Madrid: Turner Publicaciones S.L, 2013.

información en la red, la disciplina del Periodismo de Datos comenzó a desarrollarse cada vez en mayor medida.

Pero no sólo el volumen creciente de datos en la web ha impulsado esta especialización periodística; un gran impulsor y responsable del florecimiento del Periodismo de Datos es el movimiento *Open Data* y el modelo de *Gobierno Abierto*.

Primero vamos a comenzar explicando qué son los *Datos Abiertos* u *Open Data*. Éstos son datos que pueden ser utilizados, reutilizados y redistribuidos libremente por cualquier persona. La única restricción que existe es la obligación de atribuir su autoría y/o compartir los datos con su licencia de origen.

El Open Data es un movimiento que ha surgido en los últimos años, con el objetivo de poner a disposición de la sociedad datos de interés común que la conciernen. Es una “iniciativa mundial de acceso a la información del Estado por parte de los ciudadanos”⁴⁵. Esto es más específicamente lo que se conoce como *Gobierno Abierto* u *Open Government*.

Gastón Concha y Alejandra Naser señalan ocho aspectos o características sobre los cuales se basan los modelos de apertura de datos⁴⁶:

1. Los datos deben ser públicos (se deben abrir no sólo aquellos datos que las autoridades estimen pertinentes) Además, deben estar almacenados en un formato compatible con las computadoras.
2. Se deben publicar los datos originales, de la fuente primaria, sin procesamientos ni formas agregadas.
3. Los datos deben ser publicados y actualizados de forma constante y en períodos breves de tiempo para preservar su valor público.
4. Los datos deben ser accesibles para todo tipo de usuarios, sin restricción alguna.
5. Los datos se deben poder procesar por computadora.
6. Los datos deben estar disponibles para cualquier ciudadano sin necesidad de contar con un permiso o registro previo.
7. Los datos, en lo posible, deben estar disponibles en un formato en el que no exista la exclusividad de su control.

⁴⁵ Concha, Gastón, y Alejandra Naser. *Datos abiertos: Un nuevo desafío para los gobiernos de la región*. Santiago de Chile, Santiago: Naciones Unidas, 2012.

⁴⁶ Concha, Gastón, y Alejandra Naser. *Datos abiertos: Un nuevo desafío para los gobiernos de la región*. Santiago de Chile, Santiago: Naciones Unidas, 2012.

8. Los datos deben ser de uso libre, sin restricciones o sujetos a copyright, patentes o regulaciones.

El panorama en Argentina

En nuestro país el movimiento Open Data es un fenómeno relativamente nuevo. Casi no existen leyes para regular la liberación de los datos, y a los periodistas se le dificulta el trabajo a la hora de recolectar o encontrar los datos; sobre todo los relacionados con el gobierno o instituciones privadas. Es por esto que el movimiento de Gobierno Abierto es de inmensa importancia para el futuro del Periodismo de Datos.

En Argentina no existen leyes nacionales de libre acceso a la información que aseguren la liberación de los datos y el derecho de cualquier ciudadano de exigir información y transparencia por parte del Estado. La Constitución Argentina asegura el derecho de libertad de expresión y el derecho a la información, sin embargo, no hay ninguna ley específica que lo regule.

Lo que sí hay es un decreto, el decreto N°1172/03, que garantiza que el Poder Ejecutivo Nacional debe responder a los pedidos de liberación de información que sean formulados por la ciudadanía. Este decreto incluye a todos los organismos, empresas prestatarias de servicios públicos, entidades y organizaciones que reciban subsidios del Poder Ejecutivo Nacional.⁴⁷

El panorama del Open Data en la Argentina está muy lejos de ser ideal. Hay escasez de información y poca liberación de los datos. Incluso, esos pocos datos liberados usualmente son entregados en formatos no estructurados que dificultan la tarea del periodista y no coinciden con los principios de los Datos Abiertos.

⁴⁷ Perry, Felipe, y Miguel Paz. *Manual de Periodismo de Datos Iberoamericano*. 2013.

Capítulo II

LOS DATOS COMO HERRAMIENTA PERIODISTICA

Como mencionábamos anteriormente, los datos tienen la capacidad de ser la fuente del Periodismo de Datos, la herramienta con la que se narra la historia, o ambas cosas. Así es cómo éstos se convierten en la herramienta primera y fundamental de esta disciplina periodística.

El mérito del Periodismo de Datos, como dice el “Manual de Periodismo de Datos Iberoamericano”, no es tanto encontrar los datos, sino comprobar y descubrir la relación entre variables de sucesos que, a primera vista, podrían parecer aislados pero que con la implementación de las herramientas del Periodismo de Datos cobran sentido y pueden convertirse en una historia interesante.⁴⁸

“Los datos pueden revelar secretos a quienes tengan la humildad, el deseo y las herramientas para escuchar”⁴⁹, dicen Víktor Schönberger y Kenneth Cukier en el libro “Big Data, la revolución de los datos masivos”.

En esta parte del trabajo vamos a empezar a especificar cómo se obtienen esos datos, cómo se limpian y cómo se visualizan a través de la presentación de algunas de las herramientas que utiliza el Periodismo de Datos.

Cómo se utilizan y obtienen los datos

Existen diferentes formas de encontrar, analizar y visualizar los datos. Por eso es preciso mencionar las buenas prácticas y requisitos para llevar a cabo el Periodismo de Datos. El “Manual de Periodismo de Datos Iberoamericano”⁵⁰ menciona algunos de ellos:

- Datos abiertos a disposición de la audiencia (descargables o en DocumentCloud, Google Drive, etc.)
- Data viz interactiva (Tableau Public, Google Fusion Tables, etc.)
- Un mashup como cuando se cruza data de Twitter posicionando esa información en Google Maps.
- Trabajo de data mining (minería de datos).

⁴⁸ Perry, Felipe, y Miguel Paz. *Manual de Periodismo de Datos Iberoamericano*. 2013.

⁴⁹ Mayer-Schönberger, Viktor ; Cukier, Kenneth. «Big Data. La revolución de los datos masivos.» Madrid: Turner Publicaciones S.L, 2013.

⁵⁰ Perry, Felipe, y Miguel Paz. *Manual de Periodismo de Datos Iberoamericano*. 2013.

- Scraping de datos (extracción de datos automatizada).
- Uso de bases de datos propias o ajenas a partir de la cual se construye la propia.
 - Trabajo de recuperación documental web (búsqueda en internet profunda o invisible).
 - Casos de acceso a la información, pero no cuando es uno aislado, sino cuando hay un conjunto de peticiones importantes (más de diez y sistematizadas).
 - Trabajo de campo propio: es decir cuando la data no está (se basa en el análisis de webzines del gobierno, hay que construirla y crear un Excel propio).

A continuación se pueden observar algunas de las fuentes que utiliza el Periodismo de Datos:



Ilustración 5 (Gráfico propio - Fuente: “Antecedentes y fundamentos del periodismo de datos”)

La mayor parte de las veces las fuentes de información son datos públicos, como por ejemplo, los datos de las ONG, organismos del Estado, asociaciones, fundaciones, y demás. Una forma muy común también para recolectar datos es la utilización de la técnica del scraping, la cual vamos a explicar más adelante.

Una vez recolectados los datos de las fuentes antes mencionadas, es preciso seguir una serie de pasos. Sandra Crucianelli recomienda algunas prácticas útiles en Periodismo de Datos⁵¹:

- *Verificar los datos con otras fuentes*: Una vez recolectados los datos es necesario buscar más de una fuente para corroborar los mismos.
- *Chequear aleatoriamente los datos*: Sobre una muestra basada en el universo bajo estudio, es preciso chequear los datos al menos tres veces.
- *La narrativa no debe ser demasiado extensa ni sobrecargada de información numérica*: Muchos datos numéricos pueden desalentar al público a la hora de seguir leyendo la nota o producto de periodismo de datos; lo mismo sucede con la narrativa extensa.
- *Se recomienda remitir conclusiones a la visualización y mencionar el respaldo documental sobre el que se trabajó*: Una característica muy importante del periodismo de datos es la de remitir la fuente que se usó en el trabajo, los datos crudos, para que el usuario pueda verificar la información y, también, crear sus propios productos de datos.
- *Se debe hacer un uso adecuado del hipertexto*: Se recomienda compartir las fuentes tanto primarias como secundarias, pero no remitiendo a las fuentes web desde sus páginas de inicio sino a la URL permanente del sitio donde se obtuvieron los datos.
- *Se recomienda una visualización interactiva*: Que la visualización sea destacada y fuerte es un requisito fundamental para el periodismo de datos. La misma no debe utilizar una forma de recursos demasiado extensa.
- *Compartir datos abiertos con la audiencia*: Lo que puede hacerse mediante plataformas como Google Drive.
- *Mostrar a la audiencia la documentación de respaldo original completa*: Puede hacerse, por ejemplo, con Document Cloud.
- *Mantener abiertos los canales de comunicación con la audiencia*: Para una comunicación interactiva.

⁵¹ Crucianelli, Sandra. *Herramientas Digitales para Periodistas*. 2013.

- *Utilizar Tableau u otros sistemas de visualización interactiva con la opción de poder descargarlos:* Para, como se mencionó anteriormente, provocar una comunicación con la audiencia.

Cuatro productos del Periodismo de Datos

El Periodismo de Datos puede representarse de diferentes maneras, pero principalmente deviene en cuatro productos:

1. Artículos basados en datos.
2. Visualizaciones interactivas.
3. Conjunto de datos abiertos (set de datos).
4. Aplicaciones de noticias (*newsapps*).

Para explicar de qué se tratan cada uno de estos productos, es preciso retomar una vez más el libro “Herramientas digitales para periodistas” de Sandra Crucianelli, donde se especifican las características de cada uno de ellos⁵²:

1. Artículos basados en datos

Son artículos cortos, construidos a partir de grandes volúmenes de datos, como los que podrían estar contenidos en bases de datos o *datasets*.

2. Visualizaciones interactivas

Son un complemento de los artículos basados en datos. Las visualizaciones pueden también ser contenidos en sí mismos, y los artículos basados en datos, en este caso, ser un complemento.

3. Conjunto de datos abiertos (set de datos)

Es cuando se abren los datos públicos. Relacionado con el movimiento Open Data.

4. Aplicaciones de noticias

Muchas veces sucede que el volumen de datos es tan grande que resulta imposible encontrar una noticia dentro de esos datos. Por eso se diseñan aplicaciones que permiten agrupar y analizar variables (como la localización geográfica, por ejemplo) para encontrar una historia dentro de esos datos.

⁵² Crucianelli, Sandra. *Herramientas Digitales para Periodistas*. 2013.

Anteriormente mencionábamos que el producto más utilizado y más eficiente del Periodismo de Datos es la visualización interactiva, sin embargo acá podemos observar que hay distintas formas de hacer Periodismo de Datos, todas y cada una de ellas productivas y necesarias.

Estos productos de la disciplina devienen en las formas en que las audiencias interactúan con ellos. Existen cuatro niveles de lectura de datos por parte de las audiencias:

1. *Lectores esponja*: son los que buscan la esencia de la historia en un video o animación de una duración aproximada de dos minutos.
2. *Lectores habituales*: son de la “vieja escuela”, leen la noticia impresa o en línea. Lo que sería el producto “artículos basados en datos”.
3. *Lectores exploradores*: son los que interactúan con las aplicaciones interactivas y crean sus propias historias.
4. *Los topes meticulosos*: son los lectores que van directamente a la base de datos para generar sus propios cálculos y visualizaciones. Para ellos los conjuntos de datos abiertos son fundamentales⁵³.

PRINCIPALES HERRAMIENTAS UTILIZADAS

Uno de los objetivos específicos de este trabajo es el de explicar qué clase de herramientas utiliza el Periodismo de Datos para llevar a cabo su actividad.

Son cada vez más las herramientas, plataformas online y prácticas de programación que se utilizan para hacer periodismo. Éstas varían desde el uso de las herramientas de Google (Google Docs, Google Spreadsheet, etc.) y Amazon, hasta la creación de software de código abierto y la reciente aparición del movimiento Open Data.

Estas herramientas son muchas y abarcan diferentes campos de acción, por eso vamos a dividir las en tres grandes grupos: Herramientas de visualización, Herramientas de recolección y Herramientas de filtro.

⁵³ Perry, Felipe, y Miguel Paz. *Manual de Periodismo de Datos Iberoamericano*. 2013.

Herramientas de recolección

La recolección de los datos, como ya habíamos mencionado, es la primera etapa del Periodismo de Datos. Existen muchas formas para hacerse con los datos, pero la manera más fácil de lograrlo es utilizando herramientas digitales.

Antes de empezar a nombrar estas herramientas, es preciso mencionar a Paul Bradshaw y su método fundamental para encontrar datos: el *scraping de datos*.

Se define a la práctica de *scraping* como la manera de “controlar de manera automatizada sitios web con el propósito de extraer datos publicados que son de nuestro interés”⁵⁴. ¿Y cómo se controlaría de manera automatizada un sitio web? Para eso Paul Bradshaw tiene la respuesta.

Este periodista plantea que el *scraping* sería algo así como “hacer que una computadora capture información de fuentes online”⁵⁵. Para esto hay que “decirle al *scraper* qué hacer y cómo hacerlo y después sentarte y seguir haciendo las cosas que no pueden ser automatizadas”; Bradshaw aclara que este proceso “no es exactamente Wall-E”, es decir: un robot; pero que es la herramienta más poderosa que pueden utilizar los periodistas para aprehender información de la web.

En su libro “Scraping for journalists”, Paul Bradshaw explica explícita y minuciosamente cómo funciona el sistema de *scraping*, que usualmente se puede hacer de manera fácil utilizando Google Docs. A su vez, aclara que se pueden utilizar diferentes fórmulas de *scraping* para obtener diferentes resultados.

En el “Manual de periodismo de datos Iberoamericano” se recomiendan las siguientes herramientas para *web scraping*⁵⁶:

- Talented Open Studio for Data Integration and 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# (son lenguajes de programación)

⁵⁴ Perry, Felipe, y Miguel Paz. *Manual de Periodismo de Datos Iberoamericano*. 2013.

⁵⁵ Bradshaw, Paul. «Scraping for Journalists.» Leanpub, 2013.

⁵⁶ Perry, Felipe, y Miguel Paz. *Manual de Periodismo de Datos Iberoamericano*. 2013.

- Excel y Google Spreadsheets

Utilizar esta forma de recolección de datos puede parecer bastante complicado para una persona que recién se inicia en el Periodismo de Datos, pero una vez que se logran superar las barreras iniciales, esta herramienta ponderada por Paul Bradshaw es la forma más práctica y profunda de hacerse con una base de datos.

James Ball, periodista de The Guardian, dice que “la ventaja del *scraping* es que se puede hacer prácticamente con cualquier sitio, desde el pronóstico del tiempo hasta el gasto gubernamental, incluso si el sitio no tiene una API para accesos a los datos crudos”⁵⁷. Es decir, el *scraping* cumple la función del “hacker”, puede encontrar datos ahí donde no se pueden ver o donde no se quiere que se vean.

James Ball aclara que “hay límites a lo que se puede *scrapear*”, y que los factores que dificultan esta tarea pueden ser:

- Que el código HTML esté mal formateado o que incluya poco o nada de información estructural (ej: sitios oficiales más antiguos o en formatos en desuso).
- La existencia de sistemas de autenticación que se supone que impiden el acceso automático, como códigos CAPTCHA o exigencias de suscripciones pagas.
- Que los sistemas estén basados en sesiones que incluyen cookies de navegador para rastrear lo que hace el usuario.
- La falta de listados completos de ítems o la ausencia de posibilidades de búsquedas con comodines.
- El bloqueo de acceso por parte de los administradores de servidores.⁵⁸

Sin embargo, para algunos set de datos, la única manera de conseguirlos o “crearlos” es el trabajo duro, es decir: juntar archivos de Excel (cada uno conteniendo algunos registros de datos) para lograr una base de datos más amplia o completa.

Ahora bien, si bien el *scraping* de datos es la forma predilecta y más recomendada para armar una base de datos, también existen otras maneras de recolectar los datos, como puede ser, utilizando otras herramientas digitales tales como:

- Offshoreleaks.icij.org (es una base de datos en la que se basó la investigación “Panamá Papers”)

⁵⁷ «The Data Journalism Handbook.» 2011.

⁵⁸ «The Data Journalism Handbook.» 2011.

- Junar (es una plataforma de datos abiertos basada en la “nube” que está disponible de manera libre en la web)
- Tabula (es una herramienta para liberar tablas de datos que se encuentran dentro de archivos en formato PDF)
- Scrapy (fuente de datos abierta y descargable)
- Pajek (es un programa de software para Windows 32 para analizar grandes set de datos)

Las herramientas de recolección de datos son fundamentales para comenzar a trabajar en un proyecto de datos. Sin datos no hay proyecto.

Herramientas de filtro

Como Paul Bradshaw nos decía en su pirámide invertida, una vez recolectados los datos es preciso “limpiarlos”. Algunas veces, los set de datos no se encuentran en el formato deseado para poder realizar un buen trabajo de investigación, por ello es preciso convertirlos a formatos utilizables, y para eso existen varias herramientas.

Algunas de ellas son:

- Open Refine (es una herramienta que limpia los datos, los transforma de un formato a otro o los extiende con datos externos)
- Talend Open Studio (permite modelar de forma visual transformaciones de datos generando código Java)
- Google Spreadsheet (para limpiar los datos de forma manual)
- WondersharePDF (editor para pasar de PDF a texto)
- ScrapperWiki (es una plataforma online que permite descargar información de la web y agruparla de manera ordenada en una base de datos en formato Excel, CSV, etc.)
- DocumentCloud (permite extraer texto de una imagen mediante software OCR, resaltar datos, hacer anotaciones, embeber documentos, etc.)

Esta parte del proceso de un proyecto de datos es fundamental, y la utilización de herramientas digitales para llevarla a cabo es una manera práctica y recomendable para asegurar la fiabilidad de los resultados.

Herramientas de visualización

Como hemos mencionado con anterioridad, la visualización es una parte muy importante del Periodismo de Datos, ya que puede ser un producto en sí misma, y es la ventana de esta especialización periodística, la forma de publicación predilecta.

Para realizar una buena visualización, que sea comprensible para la audiencia y, sobre todo, funcional, existen varias herramientas:

- Visual.ly (para crear videos e infografías)
- Piktochart (para infografías y presentaciones imprimibles)
- Sprite (para crear gráficos de computadora)
- Many Eyes (es una herramienta de IBM para realizar infografías, gráficos de barra, gráficos de línea, etc.)
- Tiki Toki (una web basada en software para crear líneas de tiempo interactivas que pueden ser compartidas en la web)
- Datawrapper (herramienta para crear gráficos de barra, burbuja, etc. interactivos)
- Infogr.am (para crear infografías, mapas, charts o gráficos interactivos)
- iCharts (para crear gráficos de barra, burbuja, etc.)
- Dipity (para crear líneas de tiempo interactivas)
- Google Spreadsheets (donde también se pueden realizar gráficos)
- Google Maps

Dentro de las herramientas de visualización es preciso hacer un apartado para aquellas dedicadas exclusivamente a los mapas. Los mapas son fundamentales a la hora de realizar una visualización en el Periodismo de Datos, ya que son capaces de situar los hechos exactamente en el lugar indicado y presentan un servicio público, otra característica de esta disciplina.

Estas son algunas de las herramientas para ubicar hechos y personajes en un lugar exacto:

- Open Street Map (<http://www.openstreetmap.org/>) (es un proyecto colaborativo para crear mapas libres y editables. Los datos que se utilizan para realizar estos mapas son usualmente capturados con dispositivos GPS móviles)

- My Maps (<http://www.google.com/maps/d/u/0/>) (para crear mapas a través de Google Maps)
- Map Box (<http://mapbox.com/tour/>) (es un sitio web que permite crear mapas personalizados mediante software libre)
- InfoAmazonia (<http://infoamazonia.org/>) (es un sitio web que proporciona mapas ambientales de los países de la región amazónica)
- Geocommons (<http://geocommons.org/>) (es un programa para crear mapas con capas múltiples, mediante información geolocalizada de otros usuarios)
- Carto (para crear mapas interactivos con datos geoespaciales)
- Umap (permite crear mapas con capas de OpenStreetMap para su publicación online)

En esta parte del trabajo hemos hecho hincapié en solo una cantidad reducida de herramientas para recolectar, limpiar y visualizar datos. Existen muchas más, pero por cuestiones prácticas sólo hemos mencionado unas cuantas.

Capítulo III

ENTIDADES DEL PERIODISMO DE DATOS

Con la consolidación del Periodismo de Datos y la gran influencia de Internet, el uso de la tecnología se volvió un punto clave a la hora de hacer periodismo. Así fue cómo surgieron entidades que hoy en día promueven esta forma de hacer periodismo y dan cursos al respecto para capacitar a periodistas, programadores o diseñadores en la tarea del Periodismo de Datos.

A continuación vamos a nombrar a tres de ellas.

Knight Foundation

Una de las entidades más conocidas al respecto es **Knight Foundation**. Ésta es una fundación sin ánimo de lucro dedicada, en sus propias palabras, a “promover y apoyar las ideas transformacionales que promueven el periodismo de calidad, la innovación de los medios, involucran a las comunidades y fomentan las artes”⁵⁹. La Knight Foundation tiene cuatro programas: el Journalist and Media Innovation Program, el Communities Program, el Arts Program y el National Program; cada uno de los cuales utiliza una combinación de prioridades de financiación. Lo que hace esta fundación es financiar proyectos tanto periodísticos como educativos en distintas comunidades de los Estados Unidos; cuenta con *Knight Chairs* o “sillas de caballeros”, los cuales son periodistas que se encuentran en posiciones jerárquicas importantes dentro de las Universidades con programas financiados por la fundación; a su vez también los laboratorios de tecnología de periodismo son financiados en algunas universidades.

Esta fundación apoya “ideas, líderes e iniciativas que ayudan a avanzar en la práctica del periodismo, incluyendo el uso de nuevas tecnologías”⁶⁰.

La Knight Foundation puede pensarse como una fundación promotora del Periodismo de Datos o del periodismo basado en tecnología. Si bien funciona dentro del radio estadounidense, es posible pensar una organización de esta índole funcionando en países como Argentina; por qué no.

⁵⁹ *KnightFoundation*. <https://knightfoundation.org/> (último acceso: Mayo de 2017).

⁶⁰ *KnightFoundation*. <https://knightfoundation.org/> (último acceso: Mayo de 2017).

La entidad que sí tiene lugar en Argentina, en Latinoamérica y en distintas partes del mundo es **Hacks and Hackers**.

Hacks and Hackers

Hacks/Hackers (su nombre abreviado y más usado) es una organización internacional de periodismo que cuenta con decenas de “capítulos” y miles de miembros en todo el mundo, a los que cada día se agregan más. Los “capítulos” podrían definirse como los eventos que promueve esta organización, los que explicaré en un segundo. Hacks/Hackers se define a sí misma como una organización cuya misión es “crear una red de periodistas (hacks) y

“88 ciudades del mundo tienen comunidades Hacks/Hackers, que facilitan la colaboración entre periodistas y programadores” (Hidalgo, David ; Torres, Fabiola 2016).

tecnólogos (hackers) que repiensen el futuro de las noticias y la información”⁶¹. ¿Cuál es la función que cumple esta organización? Son muchas, según ellos. Puede definirse como la misión de promover y difundir conocimiento, información e ideas.

En sus “capítulos” -donde se congregan personas tan diversas como periodistas, programadores, diseñadores o simples curiosos- se realizan charlas y *hackathones*. El

hackatón es el término usado en este tipo de comunidades de hackers para referirse al encuentro de programadores, periodistas, diseñadores, etc. donde se desarrollan proyectos en conjunto de los cuales surgen tanto productos de Periodismo de Datos como también desarrollos de software libre.

Esta organización también ofrece cursos de codificación para principiantes para el uso de herramientas como Google Fusion Tables y Js.

Los “capítulos” de Hacks/Hackers se organizan alrededor del mundo, tantas veces sean necesarios, dependiendo de la demanda de cada zona. En Argentina “Hacks/Hackers (Buenos Aires)” es el capítulo más importante, pero también existe en otras provincias del país, como mencionaremos más adelante.

La organización, en su sitio web oficial, se define a sí misma como “una comunidad digital de personas que buscan inspirarse mutuamente, compartir información (y código) y

⁶¹ *Hacks and Hackers*. <https://hackshackers.com/> (último acceso: Mayo de 2017).

colaborar para inventar el futuro de los medios de comunicación y del periodismo”⁶², asegura que su objetivo es reunir a todas las personas que están ayudando para dar sentido al mundo, que la organización “es para los hackers que exploran las tecnologías para filtrar y visualizar información, y para los periodistas que utilizan la tecnología para encontrar y contar historias”.

En definitiva, Hacks and Hackers trabaja por y para este grupo de gente que se dedica a hacer Periodismo de Datos, a innovar en la disciplina y a encontrar nuevas historias basadas en datos.

Wikileaks

Si estamos hablando de entidades significativas en el universo del Periodismo de Datos, es preciso mencionar una que dio comienzo a este movimiento: **Wikileaks**.

Todos alguna vez habrán oído hablar de Wikileaks, sobre todo a partir de 2010.

Wikileaks es una organización de medios internacional sin ánimos de lucro que a su vez también cumple la función de “biblioteca asociada”. Fue fundada por Julian Assange en 2006 y se especializa en “el análisis y publicación de grandes conjuntos de datos de

“La manera en que organizaciones de noticias como The Guardian y el New York Times manejaron las grandes cantidades de datos difundidos por Wikileaks, es uno de los grandes casos que impulsaron el término de Periodismo de Datos” (The Data Journalism Handbook 2011).

materiales oficiales censurados o restringidos que involucran temas como la guerra, el espionaje y la corrupción”⁶³ según describe en su página web oficial. Esta organización, hasta la fecha, ha filtrado más de 10 millones de documentos relacionados con estos temas.

“Wikileaks es una biblioteca gigante de los documentos más perseguidos del mundo. Damos asilo a estos documentos, los analizamos, los promovemos y obtenemos más”⁶⁴ dijo su fundador.

Esta organización, publica, a través de su sitio web, informes anónimos y documentos filtrados. Y al tener estos documentos un cariz “oculto” o

⁶² *Hacks and Hackers*. <https://hackshackers.com/> (último acceso: Mayo de 2017).

⁶³ *Wikileaks*. (<https://wikileaks.org/>) (último acceso: Mayo de 2017).

⁶⁴ *Wikileaks*. (<https://wikileaks.org/>) (último acceso: Mayo de 2017).

“clasificado”, Wikileaks asegura tener “relaciones contractuales y vías de comunicación seguras con más de 100 grandes organizaciones de medios de todo el mundo”⁶⁵ y recalca que “esto le da a las fuentes de Wikileaks recursos para negociar el poder, el impacto y las protecciones técnicas que de otro modo serían difíciles o imposibles de lograr”.

Esta organización, debido al contenido delicado de la información y archivos de datos que brinda, ha sufrido varios intentos de censura, ya sea por parte de gobiernos como por parte de otras organizaciones que se vieron amenazadas por la competencia. Sin embargo, Wikileaks asegura mantener una “resistencia a todos los intentos de censura”, y ha logrado superarlos en su totalidad.

La filosofía de Wikileaks es la de preservar el anonimato de las fuentes que realizan las filtraciones, garantizando que el contenido circule y se propague. Pablo Mancini dice que “Wikileaks mostró en muy poco tiempo que en el caos de la red y la información a escala, el periodismo puede ir mucho más allá que usar la tecnología sólo para hacer lo mismo de siempre pero en tiempo real”⁶⁶. Esta organización abrió las puertas para que el periodismo y la tecnología, en conjunto, puedan ser vistos en una primera instancia, y luego utilizados, como una herramienta para hacer Periodismo de Datos, de contar las historias de una manera nueva e innovadora.

Pablo Mancini describe cinco características fundamentales que hacen de Wikileaks una organización confiable para practicar el Periodismo de Datos⁶⁷:

1. El medio está en todas partes, no solo en un país.
2. Se tiene el derecho a reservar la fuente y a encriptarla.
3. Los *publishers* tradicionales (es decir: los diarios tradicionales) cumplen el papel de amplificadores de su contenido generado por fuera de los canales tradicionales.
4. La descarga, la copia y la distribución del contenido está al servicio de garantizar el acceso a la información.
5. La filtración sin inteligencia es invisible a los ojos de la audiencia, por lo que Wikileaks plantea la relevancia de los análisis por sobre la información bruta, lo que siempre fue la base del periodismo.

⁶⁵ Wikileaks. (<https://wikileaks.org/> (último acceso: Mayo de 2017).

⁶⁶ Mancini, Pablo. *Hackear el periodismo*. Buenos Aires: La Crujía Ediciones, 2011.

⁶⁷ Mancini, Pablo. *Hackear el periodismo*. Buenos Aires: La Crujía Ediciones, 2011.

PIONEROS DEL PERIODISMO DE DATOS

Una vez definidas las principales entidades que componen el Periodismo de Datos, es menester mencionar los casos en que esta disciplina fue aplicada en los medios de comunicación. Para esta parte del trabajo he seleccionado dos ejemplos, uno a nivel mundial y otro a nivel nacional. El fin de este apartado es ejemplificar casos de Periodismo de Datos en los medios para poder alcanzar una mayor comprensión de qué es y qué significa esta disciplina.

En el mundo: El caso The Guardian

El diario inglés The Guardian es uno de los pioneros en el Periodismo de Datos, principalmente porque a partir de 2009 inició un *datablog*, desde el cual realiza investigaciones basadas en datos y publica sus *datasets* y datos en crudo, que están abiertos a disposición de la audiencia.

Gracias a su Datablog, The Guardian se ha convertido en un modelo a seguir para otros periódicos online.

El equipo de Datablog está formado por cinco profesionales, según explica la Dra. María Ángeles Chaparro Domínguez: Simon Rogers (editor del departamento), Ami Sedghi (investigadora de datos y reportera), John Burn-Murdoch (investigador de datos y reportero), James Ball (periodista de datos) y Katy Stoddard (periodista de investigación). En este mismo texto, Chaparro Domínguez explica que en sus cuatro años de vida, el Datablog de The Guardian “ha publicado numerosos reportajes escritos, que en la mayoría de los casos se han completado con mapas y otras herramientas interactivas”⁶⁸ e introduce el ejemplo del reportaje publicado por el Datablog en agosto de 2012 bajo el título “Could you be an athlete? Olympics 2012 by age, weight and height”, donde se analizaron datos de los 10.383 deportistas que participaron en las Olimpiadas de Londres en ese año.

Este es un ejemplo claro de Periodismo de Datos, donde los datos representan una suma enorme de información que es analizada minuciosamente, a la que se le hacen diferentes preguntas y es publicada mediante visualizaciones y textos basados en datos.

⁶⁸ Chaparro Domínguez, María Ángeles. *La evolución del periodismo de precisión: el blog de The Guardian sobre periodismo de datos*. Logroño: Universidad Internacional de La Rioja, 2013.

El editor de este Datablog, Simon Rogers, explica que cuando su equipo y él lanzaron el blog, no tenían “idea a quién podrían interesarle los datos en crudo, las estadísticas y visualizaciones”⁶⁹, algo que le pasó a muchos de los pioneros en esta especialización.

El Datablog de The Guardian cumple varias funciones, que varían desde búsquedas de datos de gobiernos y desarrollo global, visualizaciones de datos realizadas por artistas gráficos del mismo periódico y de toda la red, hasta la creación de herramientas para explorar datos de gasto público.

Rogers dice que todos los días utilizan “plantillas de cálculos de Google para compartir los datos completos que respaldan nuestro trabajo; visualizamos y analizamos esos datos y luego los usamos para proveer historias para el diario y el sitio”⁷⁰. Rogers dice haberse sorprendido ante la respuesta de los lectores que se mostraban cada vez más interesados en los datos en crudo detrás de las noticias, y agrega: “Cuando lanzamos el Datablog, creíamos que el público serían programadores buscando crear aplicaciones. De hecho es gente que quiere saber más sobre las emisoras de carbono, inmigración de Europa oriental, el desglose de las muertes en Afganistán, o incluso la cantidad de veces que los Beatles usaron la palabra ‘amor’ en sus canciones”⁷¹; que, de hecho, fueron 613 veces.

Lo que cuenta este periodista se debe, en una primera instancia, al cambio de paradigma que sufrió el periodismo con movilizaciones tales como el Open Data, Gobierno Abierto o Wikileaks. El campo del periodismo cambió, al igual que su audiencia, y ahora el lector busca objetividad en los datos, quiere ir directo a la fuente.

Simon Rogers cree que “el evento que cambió el juego para el periodismo de datos se dio en la primavera de 2010, comenzando por una hoja de cálculo de 92.201 filas de datos, cada una conteniendo un desglose detallado de un evento militar en Afganistán”⁷². A lo que se refiere Rogers son a los registros de guerra que Wikileaks filtró en ese mismo año, los que revolucionaron al mundo del Periodismo de Datos.

Pero no sólo Wikileaks ayudó a cambiar la panorámica de esta especialización periodística, el periódico The Guardian, con su Datablog, ha contribuido a crear una forma de hacer periodismo que todavía tiene mucho futuro por delante, que se encuentra, recién ahora, en sus primeros pasos. Tal es la importancia de este Datablog que otros medios de

⁶⁹ Perry, Felipe, y Miguel Paz. *Manual de Periodismo de Datos Iberoamericano*. 2013.

⁷⁰ Perry, Felipe, y Miguel Paz. *Manual de Periodismo de Datos Iberoamericano*. 2013.

⁷¹ Perry, Felipe, y Miguel Paz. *Manual de Periodismo de Datos Iberoamericano*. 2013.

⁷² Perry, Felipe, y Miguel Paz. *Manual de Periodismo de Datos Iberoamericano*. 2013.

comunicación alrededor del mundo han tomado su ejemplo para introducir el Periodismo de Datos en sus propias redacciones. Ese es el caso, por ejemplo, de la Nación Data.

En Argentina: El caso La Nación Data

La iniciativa de realizar una investigación basada en datos y de incursionar en ese panorama tan nuevo e indómito que era el Periodismo de Datos surgió, para el diario La Nación de Argentina, a partir de las reducciones de subsidios al sistema de transporte público de pasajeros en 2011.

Hasta ese momento no existía en nuestro país un medio de comunicación que utilizara los datos de la manera en que lo hace esta disciplina. Se desconocía qué era el Periodismo de Datos y su aplicación a las salas de redacción.

Angélica Peralta Ramos, periodista de La Nación, explica en el “Manual de Periodismo de Datos 1.0” la experiencia de comenzar a hacer Periodismo de Datos.

“Desde 2002 los subsidios para el sistema de transporte público de pasajeros en la Argentina han estado creciendo de modo exponencial, rompiendo un record cada año. Pero en 2011, luego de ganar las elecciones, el nuevo gobierno argentino anunció reducciones de los subsidios para los servicios públicos a partir del mes de diciembre de ese año. Al mismo tiempo, decidió transferir la administración de líneas locales de ómnibus y del subte al Gobierno de la Ciudad de Buenos Aires” dice Peralta Ramos, y agrega que “el Gobierno porteño rechazó esta decisión”⁷³. Durante este panorama fue que Angélica Peralta Ramos y sus colegas del diario se reunieron por primera vez a discutir cómo iniciar una propia investigación de Periodismo de Datos, basándose en las grandes investigaciones del diario The Guardian.

Esta periodista asegura que el éxito de ese primer proyecto de periodismo basado en datos ayudó al diario internamente para comenzar a pensar en la creación de una sección de noticias hechas con Periodismo de Datos, una sección que cubra el periodismo de investigación y que a su vez provea servicio al público. Los resultados de este planteamiento se vieron materializados en una plataforma donde el diario publica datos abiertos sobre distintos tópicos de interés públicos, en formatos procesables por computadora, siguiendo las nociones del Open Data. La plataforma es (<http://data.lanacion.com.ar/>)

⁷³ «The Data Journalism Handbook.» 2011.

El diario La Nación no solo cuenta con esta plataforma de publicación de datos abiertos, sino que también, inspirándose en el Datablog de The Guardian, cuenta con un blog que cumple funciones similares a éste. Según sus propias palabras, “la propuesta de NACIÓN DATA es registrar y compartir casos de periodismo de bases de datos, experiencias que llevan años de trabajo como también iniciativas recientes en las que se trabaja con la creación de visualizaciones de datos, la interacción y la participación del usuario. También nos ocuparemos de difundir aquellos espacios en los que trabajan programadores y periodistas de manera conjunta para resolver un tema específico, el movimiento mundial hacia apertura de datos públicos y gobiernos 2.0, y las diferentes herramientas y tecnologías de bases de datos, para comprender qué implica gestionar y analizar datos, y cómo se enriquece la disciplina cuando otras ciencias agregan valor a los mismos”⁷⁴. A su vez, este blog incluye los casos, bases de datos y visualizaciones interactivas publicados en el diario.

Como se puede ver, La Nación Data es un blog que cumple con todas las características fundamentales del Periodismo de Datos: utilización de bases de datos para generar contenido, liberación de los datos, publicación de los datos crudos, visualizaciones interactivas, etc.

Y acá no termina. El diario La Nación cuenta con una plataforma de liberación de datos, un blog como el de The Guardian y una sección exclusiva sobre investigaciones basadas en datos en su diario online.

En esta sección del diario, La Nación publica sus contenidos específicos de Periodismo de Datos, ya sean noticias basadas en datos, noticias sobre becas y estudios de Periodismo de Datos, y visualizaciones interactivas, entre otras cosas. También cuenta con un acceso directo a su base de datos abierta que se llama *VozData* la cual es “una plataforma de colaboración abierta creada para transformar documentos públicos en información útil. Es también una iniciativa para amplificar la voz de los datos que de otra manera permanecen silenciosos y distantes del control o la participación ciudadana”⁷⁵.

En definitiva, el producto de Periodismo de Datos que ofrece el diario La Nación de Argentina es uno de los más completos del mundo, comparable con grandes proyectos como el Datablog de The Guardian. Y se puede asegurar que, así como el Datablog fue pionero en el Periodismo de Datos en Inglaterra, La Nación Data fue pionero en el Periodismo de Datos en Latinoamérica.

⁷⁴ *La Nación Data*. <http://data.lanacion.com.ar/> (último acceso: Mayo de 2017).

⁷⁵ *La Nación Data*. <http://data.lanacion.com.ar/> (último acceso: Mayo de 2017).

Capítulo IV

ALGUNOS EJEMPLOS DE PERIODISMO DE DATOS

Ahora llegó el momento de ver más directamente cómo trabaja el Periodismo de Datos. Para entender un poco más de qué se trata el Periodismo de Datos y qué clase de investigaciones puede producir, es preciso analizar algunos ejemplos de investigaciones basadas en datos. Para ello vamos a retomar a los tres referentes de esta disciplina mencionados anteriormente: The Guardian, La Nación y Wikileaks.

En esta parte del trabajo final de carrera me he basado en los ejemplos de Periodismo de Datos publicados en el libro “The Data Journalism Handbook”. Así que a continuación, todas las referencias textuales pertenecen a este manual, exceptuando algunas especificaciones de contenido que he observado por mi cuenta.

Datblog de The Guardian

Un ejemplo de investigación de datos realizada por el diario inglés The Guardian, publicada en su Datblog, es expuesta en “The Data Journalism Handbook”. Se trata de la cobertura por parte de este diario de los disturbios ocurridos en 2011 en Reino Unido.

Durante el verano de 2011, en Reino Unido, se produjo una oleada de disturbios, asociados, según la opinión de algunos políticos, con acciones no vinculadas con la pobreza; estos políticos aseguraban que lo saqueos fueron simplemente criminales. A su vez, el primer ministro de Reino Unido, junto con otros políticos conservadores, echó la culpa a los medios sociales y alegaron que los disturbios fueron organizados e incitados a través de las redes sociales.

En este contexto y por este contexto, The Guardian (en colaboración con la London School of Economics) creó un proyecto de Periodismo de Datos para abordar estas cuestiones, que se llamó **Reading the Riots**

(<https://www.theguardian.com/uk/series/reading-the-riots>).

En “The Data Journalism Handbook” se explica que “el diario usó Periodismo de Datos a gran escala para permitir al público comprender mejor quién saqueaba y por qué. También trabajaron con otro equipo de académicos, encabezados por el profesor Rob Procter de la Universidad de Manchester para entender mejor el rol de los medios sociales

(redes sociales), que The Guardian mismo había usado abundantemente para sus informes durante los disturbios”⁷⁶ (The Data Journalism Handbook 2011).

El equipo de trabajo de este proyecto fue liderado por Paul Lewis, el Editor de Proyectos Especiales de The Guardian, quien cubrió los disturbios por su cuenta de Twitter.

Twitter fue la plataforma en la que se basó The Guardian para recabar información. El diario “trabajó a partir de 26.000.000 de tweets sobre los disturbios. El objetivo principal de este trabajo con los medios sociales fue ver cómo circulan los rumores en esa red social, la función que tienen distintos usuarios/actores en la propagación y difusión de flujos de información, para ver si se usó la plataforma para incitar, y para examinar otras formas de organización”.

En “The Data Journalism Handbook” se divide al proyecto *Reading the Riots* en dos fases: la primera es “el período de los disturbios mismos y las maneras en que los datos ayudaron a narrar historias mientras se desarrollaban los disturbios”; la segunda es “la investigación mucho más intensa con dos conjuntos de equipos académicos trabajando con The Guardian, para recolectar datos, analizarlos, y escribir informes con análisis de fondo sobre lo descubierto”⁷⁷.

Los resultados de la primera fase (que representa el momento actual en el que estaban sucediendo los disturbios) fueron publicados enteramente durante una semana de cobertura a comienzos de diciembre de 2011. Este es un aspecto que caracteriza al Periodismo de Datos; en esta disciplina, al trabajar con grandes volúmenes de información, es posible realizar una serie de publicaciones, reutilizar los datos para completar investigaciones pasadas o futuras, y seguir el hilo de la historia mediante visualizaciones que se renuevan. Así fue cómo decidió contar The Guardian su historia: mediante una serie de publicaciones que incluían desde visualizaciones interactivas hasta análisis profundos volcados en textos.

A continuación se exponen algunos ejemplos claves de cómo el diario The Guardian utilizó el Periodismo de Datos en el proyecto *Reading the Riots*.

Primera fase

Durante la primera fase -en el momento en que los disturbios estaban sucediendo- el diario utilizó mapas simples donde se mostraron las localizaciones de los lugares en donde estaban ocurriendo los disturbios, combinando datos de pobreza con dichos lugares, para

⁷⁶ «The Data Journalism Handbook.» 2011.

⁷⁷ «The Data Journalism Handbook.» 2011.

desmentir la opinión de los políticos que argumentaban que no existía un vínculo entre los *rioters* y la pobreza. Para lograr esto se utilizaron herramientas de mapeo preexistentes.

A continuación se pueden ver algunas imágenes de ejemplo.

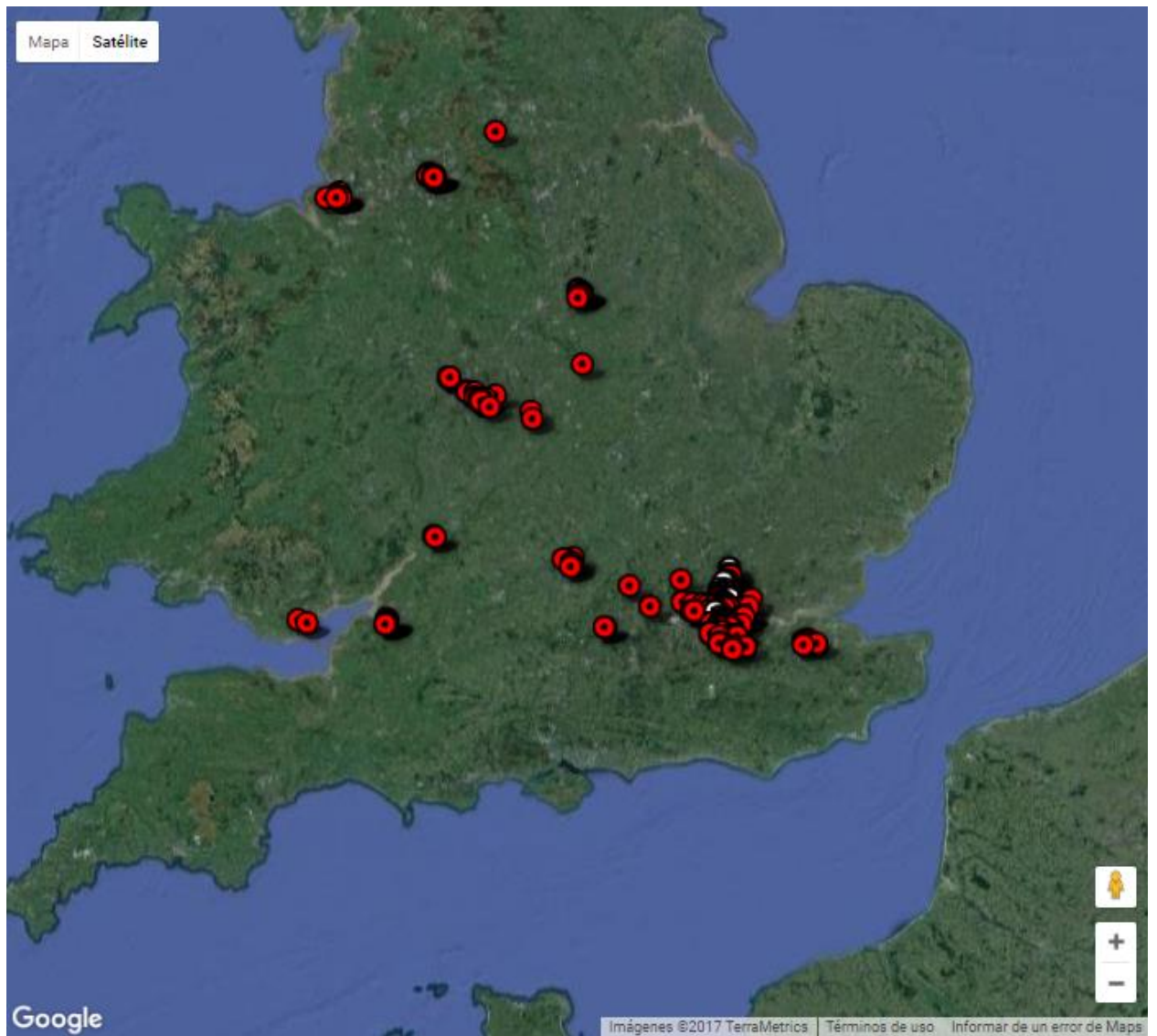


Ilustración 6 (Fuente: <https://www.theguardian.com/news/datablog/interactive/2011/aug/09/uk-riots-incident-map>)

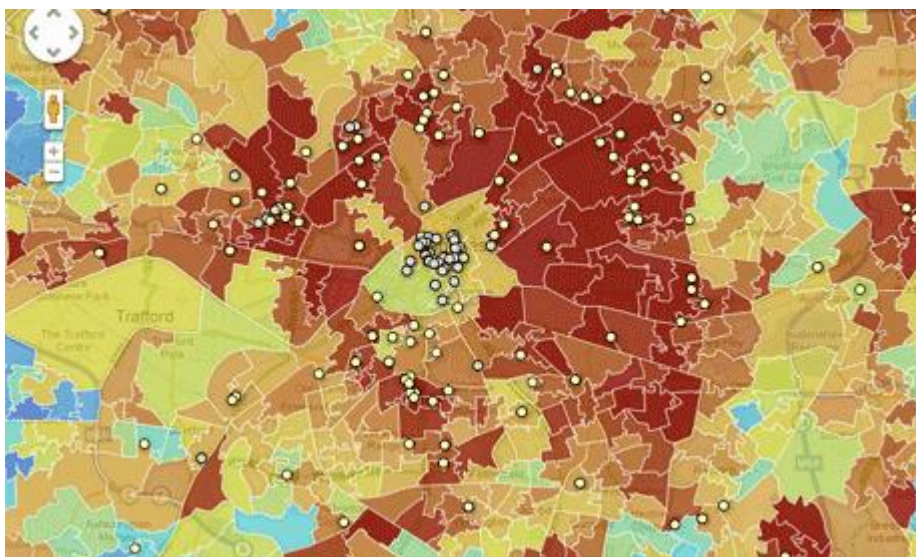


Ilustración 7 (Fuente: <https://www.theguardian.com/news/datablog/2011/aug/16/riots-poverty-map-suspects>)

En el primer mapa los disturbios se visualizan en las zonas satelitales correspondientes a Reino Unido. El segundo mapa es una especie de “mapa de calor” donde se muestran las zonas de mayor conglomración de *rioters*; el rojo representa la mayor cantidad y el verde la menor.

Durante esta fase se pudo comprobar que “en relación al uso de medios sociales durante los disturbios (en este caso, Twitter), el diario creó una visualización de *hashtags*, relacionadas con los disturbios usadas durante este período, lo que destacó que Twitter fue usado principalmente para responder a disturbios en vez de para organizar a gente para saquear”⁷⁸.

Segunda fase

En esta segunda fase se realizó un análisis de los disturbios. The Guardian informó sus conclusiones dos meses después de los sucesos, luego de haber realizado investigaciones intensivas.

Para exponer las conclusiones de su investigación, el Datablog realizó dos visualizaciones importantes. Una de ellas es un video corto (<https://www.theguardian.com/uk/datablog/video/2011/dec/05/england-riots-commute-map>) donde se muestran los resultados de la primera fase donde se combinaron las localizaciones de los disturbios con los lugares de procedencia de los *rioters*, mostrando esta relación con lo que se llamó *Viaje a los disturbios*. “Aquí el diario trabajó con un

⁷⁸ «The Data Journalism Handbook.» 2011.

especialista en mapeo de transporte, ITO World, para hacer un modelo de la ruta más probable utilizada por quienes protagonizaron los disturbios al dirigirse a los distintos lugares donde saquearon, lo que destaca patrones diferentes para distintas ciudades, con viajes largos en algunas de ellas”⁷⁹.

La segunda visualización realizada para este proyecto hace referencia a las maneras en las que se extienden los rumores en Twitter. Para esto se analizaron siete rumores surgidos durante los disturbios, propagados por esta red social. Así fue cómo el equipo académico de The Guardian “recolectó todos los datos relacionados con cada rumor y diseñó un código que identificara cada tweet de acuerdo a los cuatro códigos principales: gente que simplemente repite el rumor (afirma algo), lo rechaza (afirma algo contrario), lo cuestiona (interrogación) o simplemente lo comenta (comentario). Todos los tweets fueron codificados por triplicado y los resultados fueron visualizados por el equipo interactivo de The Guardian”⁸⁰.

A continuación se pueden observar estas visualizaciones.

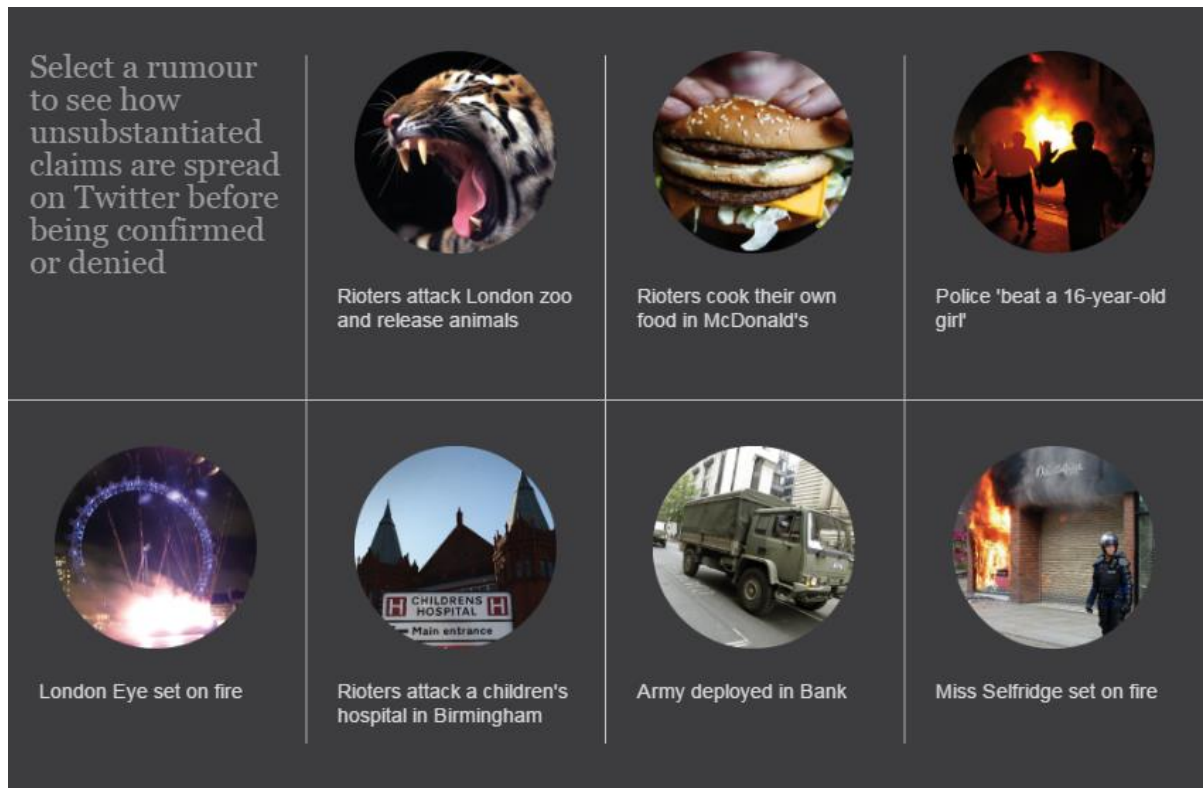


Ilustración 8 (Fuente: <https://www.theguardian.com/uk/interactive/2011/dec/07/london-riots-twitter>)

⁷⁹ «The Data Journalism Handbook.» 2011.

⁸⁰ «The Data Journalism Handbook.» 2011.

Como se puede ver, algunos de los rumores propagados por Twitter eran realmente increíbles. Tales como:

- Los *rioters* atacaron el Zoológico de Londres y liberaron a los animales.
- Los *rioters* cocinaron su propia comida en McDonald's.
- La policía golpeó a una chica de 16 años.
- El London Eye fue prendido fuego.
- Los *rioters* atacaron un hospital de niños en Birmingham.
- El ejército se desplegó en un banco.
- Se prendió fuego Miss Selfridge.

El usuario, en esta visualización interactiva, puede seleccionar un rumor y ver cómo éste sigue un proceso de negación o confirmación en Twitter antes de que las fuentes mismas puedan explicar la situación.

En la imagen siguiente se puede observar cómo funciona la aplicación.



Ilustración 9 (Fuente: <https://www.theguardian.com/uk/interactive/2011/dec/07/london-riots-twitter>)

Las burbujas verdes representan la supuesta “confirmación” del rumor, las burbujas naranjas representan la “pregunta/duda” de ese rumor, y las grises cuando el tweet sólo

hace referencia a un comentario. En la línea del tiempo que aparece por encima de las burbujas se puede ir variando la franja horaria desde que comenzó el rumor hasta que se confirmó, y se puede comprobar que las burbujas van cambiando de tamaño y color, finalizando en la confirmación o negación del humor de forma oficial. A la izquierda de la visualización se pueden leer los tweets más influenciados que conforman las burbujas.

El “The Data Journalism Handbook” expone que “el rol de los principales medios es evidente en algunos de estos rumores (por ejemplo, rechazándolos abiertamente, o confirmándolos rápidamente como noticias), al igual que la naturaleza correctiva de Twitter mismo en términos de responder a tales rumores”⁸¹.

Este estudio de caso del proyecto *Reading the Riots* de el Datablog de The Guardian es un ejemplo perfecto para explicar el alcance que puede tener una investigación periodística basada en datos. Este proyecto muestra cómo los datos pueden ser utilizados para narrar historias importantes, y por sobre todo: para darle al usuario una investigación basada en datos verificables y concretos, que ellos mismos pueden utilizar para contar otras historias o hacer otro análisis de los hechos. En este caso el usuario podía comprobar la veracidad o falsedad de los rumores sobre los disturbios en Reino Unido.

La Nación Data

Otro ejemplo para analizar de Periodismo de Datos es una investigación realizada por el diario La Nación de Argentina, donde se analizaron los cambios en los subsidios a colectivos después de las elecciones del año 2011.

Esta investigación fue expuesta y analizada por “The Data Journalism Handbook”, el cual expresa que “en 2011, luego de ganar las elecciones, el nuevo gobierno argentino anunció reducciones de los subsidios para los servicios públicos a partir del mes de diciembre de ese año. Al mismo tiempo, decidió transferir la administración de líneas locales de ómnibus y del subte al Gobierno de la Ciudad de Buenos Aires”⁸².

En este contexto fue que Angélica Peralta Ramos, periodista de La Nación, junto con sus colegas del diario, se reunieron por primera vez a discutir una iniciativa de Periodismo de Datos en el diario, como se mencionó anteriormente en esta tesis.

Peralta Ramos explica, en “The Data Journalism Handbook” que “las malas condiciones del sistema de transporte público afectan la vida de más de 5800000 pasajeros diarios.

⁸¹ «The Data Journalism Handbook.» 2011.

⁸² «The Data Journalism Handbook.» 2011.

Demoras, huelgas, desperfectos de vehículos, o incluso accidentes suceden a menudo”⁸³. Es por esto que el equipo recientemente formado de Periodismo de Datos de el diario La Nación, decidió analizar a dónde se dirigen los subsidios para el sistema de transporte público en la Argentina, para poner esos datos a disposición de los ciudadanos por medio de una aplicación de noticias realizada por La Nación Data llamada **Explorador de Subsidios del Transporte**⁸⁴.



Ilustración 10 (Fuente: <http://www.lanacion.com.ar/subsidios-a-colectivos-t48385>)

Para realizar esta investigación, el equipo de La Nación Data comenzó calculando cuánto reciben por mes las compañías de ómnibus por parte del Estado. Para hacerlo, tomaron los datos disponibles en el sitio web del Departamento de Transporte, “donde se publican desde 2006 más de 400 PDF conteniendo pagos mensuales en efectivo a más de 1300 compañías”⁸⁵.

Al momento de realizar esta tesis, el sitio web del Departamento de Transporte se encuentra deshabilitado, por lo que no se puede comprobar si los datos expuestos en este trabajo son exactos.

Para recopilar toda esa información, el equipo de La Nación incorporó a un programador “para desarrollar un recopilador de información de modo de automatizar la descarga regular y la conversión de los PDF en archivos de Excel y bases de datos”. Estas bases de datos pueden encontrarse en el diario online, en el apartado “La Nación Data”, donde se puede acceder a los datos crudos de la investigación.

⁸³ «The Data Journalism Handbook.» 2011.

⁸⁴ *La Nación Data*. <http://data.lanacion.com.ar/> (último acceso: Mayo de 2017).

⁸⁵ *La Nación Data*. <http://data.lanacion.com.ar/> (último acceso: Mayo de 2017).

El siguiente paso de la investigación “fue identificar cuánto le cuesta en promedio al gobierno el mantenimiento mensual de un vehículo de transporte público”⁸⁶. Para saber esto se consultó a otro sitio web oficial, el de la Comisión Nacional de Regulación de Transporte (que al momento de realizar esta tesis se encuentra deshabilitado). Esta comisión se encarga de regular el transporte en la Argentina. Y lo que encontró el equipo de La Nación Data en el sitio web de la comisión fue “una lista de compañías de ómnibus que poseen en total 9000 vehículos”⁸⁷ (a la fecha de la investigación). Para analizar estos datos, el equipo de periodistas y programadores desarrolló un normalizador, para poder conciliar los nombres de las compañías de transporte y poder hacer referencias cruzadas entre dos conjuntos de datos: la cantidad de vehículos y los nombre de las compañías.

Lo que se hizo a continuación fue buscar el número de registro de cada vehículo en el sitio de la CNRT. “En Argentina, estos registros están compuestos de letras y cifras que se corresponden con la edad del vehículo”⁸⁸. A partir de ahí se hizo el cálculo inverso para descubrir la edad promedio de los ómnibus y mostrar cuánto dinero recibe cada compañía, para después comparar los montos en base a la edad promedio del vehículo.

En “The Data Journalism Handbook” se menciona que “en medio de este proceso, cambió misteriosamente el contenido de los PDF oficiales con los datos, aunque las URL y los nombre de los archivos se modificaron. En algunos PDF faltaban los ‘totales’ verticales, lo que hace imposible cruzar los mismos en todo el período investigado”⁸⁹ que va de 2002 a 2011.

En este momento, como mencionaba anteriormente, esos datos ya ni siquiera existen, o por lo menos no se encuentran disponibles en el sitio web, ya que el mismo está fuera de funcionamiento.

El siguiente paso que tomó el equipo de La Nación Data para continuar con la investigación, fue llevar el proyecto a “un hackatón organizado por Hacks/Hackers en Boston, donde el programador Matt Perry generosamente creó lo que se llama el ‘Espía de PDF’ (...) El espía de PDFs apunta a una página web llena de PDF y verifica si el contenido dentro de los PDF ha cambiado”⁹⁰.

⁸⁶ «The Data Journalism Handbook.» 2011.

⁸⁷ «The Data Journalism Handbook.» 2011.

⁸⁸«The Data Journalism Handbook.» 2011.

⁸⁹ «The Data Journalism Handbook.» 2011.

⁹⁰ «The Data Journalism Handbook.» 2011.

Así fue cómo se reveló un intento por parte del Gobierno de bloquear esos datos para que no fueran analizados.

En esta parte de la investigación surge una problemática que en Argentina y Latinoamérica es muy recurrente: la restricción o modificación de los datos públicos. Hecho que justifica una vez más la necesidad de transparencia, de la implementación de un Gobierno Abierto y del movimiento Open Data.

La visualización

Para exponer esta investigación, el equipo de La Nación Data realizó varios productos de Periodismo de Datos, que variaban desde aplicaciones de noticias (el proyecto en sí mismo), bases de datos abiertas (publicadas en la *app* de noticias), noticias basadas en datos (publicadas en la *app* de noticias y en el diario online) y visualizaciones interactivas.

Estos son algunos ejemplos de las visualizaciones:

PROGRESIÓN DE SUBSIDIOS POR MES

Cuánto se destina mensualmente a las líneas de colectivos. Cada cifra corresponde a la suma del dinero en efectivo y el gasoil. (Actualizado a febrero de 2016 - Archivos procesados 475 - Procesado 25/04/2016).

Gráfico mensual



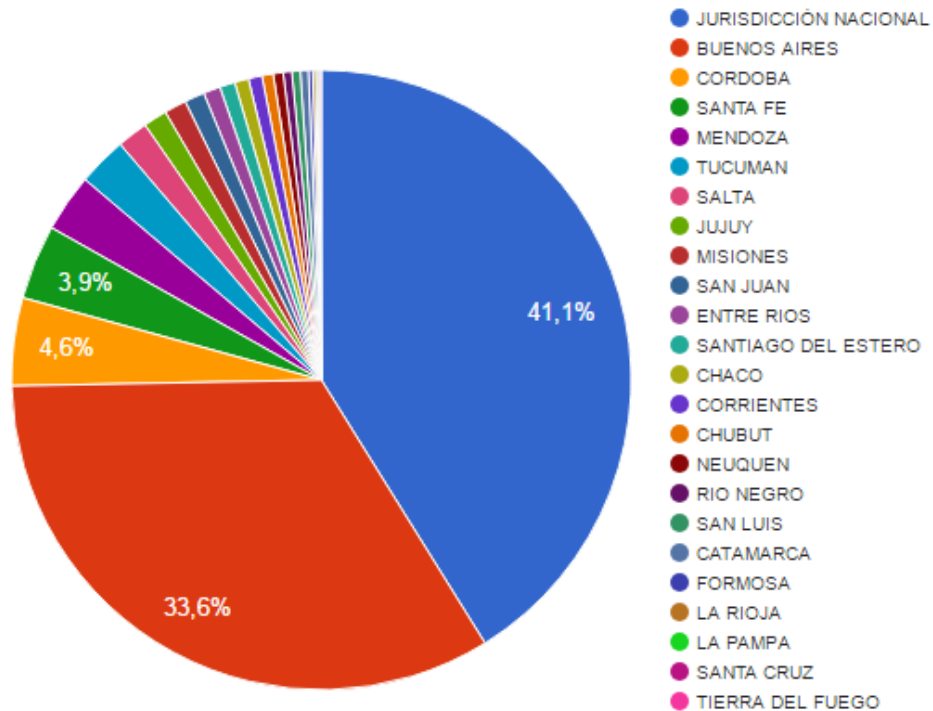
Ilustración 11 (Fuente: <http://interactivos.lanacion.com.ar/Intransporte/charts/1>)

En esta visualización se pueden observar los montos que se destinan mensualmente a las líneas de colectivos, y su progresión por año. Las cifras corresponden a la suma del dinero

en efectivo y gasoil. En este caso el usuario puede ir moviéndose en el gráfico para encontrar datos más específicos.

RANKING DE JURISDICCIONES SUBSIDIADAS

Comparación de las distintas jurisdicciones. No incluye el gasoil porque el mismo no está discriminado por jurisdicción. (Actualizado a febrero de 2016 - Archivos procesados 475 - Procesado 25/04/2016).



Nota: Jurisdicción Nacional corresponde a las líneas de colectivos que circulan por la Ciudad de Buenos Aires exclusivamente, o tocan la Capital en algún punto de su recorrido.

Ilustración 12 (Fuente: <http://interactivos.lanacion.com.ar/Intransporte/charts/4>)

En esta visualización se muestra el ranking de las jurisdicciones subsidiadas; cada color del gráfico de torta representa una provincia, y si el usuario pasa el mouse por encima de cada zona del gráfico puede ver el porcentaje específico de la jurisdicción.

Angélica Peralta Ramos asegura que el proyecto tuvo un gran impacto, y que la aplicación web realizada por su equipo ha tenido decenas de miles de visitas, e incluso la investigación apareció en la primera plana de la edición impresa del diario.

El equipo de La Nación Data pudo demostrar cómo una investigación basada en datos puede llegar a ser una fuente importante de información que puede devenir en distintos beneficios para los medios de comunicación. Este equipo realizó una investigación similar

a la de The Guardian, ya que volcó la información en distintos productos de Periodismo de Datos, siguiendo el hilo de la historia mediante varias vías de visualización.

Hacks/Hackers capítulo Buenos Aires: Electoral Hack

El ejemplo que vamos a presentar ahora es una aplicación de noticias llamada **Electoral Hack**, cuya función fue mostrar los resultados, a medida que se iban generando, de las elecciones presidenciales en Argentina en 2011.

“The Data Journalism Handbook” explica que “*Electoral Hack* es un proyecto de análisis político que visualiza datos de los resultados provisionales de las elecciones de octubre de 2011 en la Argentina. El sistema también incluye información de anteriores elecciones y estadísticas demográficas de todo el país. El proyecto fue actualizado en tiempo real con información del recuento provisional de las elecciones nacionales de 2011 en el país y dio resúmenes de los resultados”⁹¹.

Este proyecto surgió como una iniciativa del capítulo de Hacks/Hackers Buenos Aires, en un esfuerzo colaborativo de periodistas, programadores, diseñadores, analistas, científicos políticos, etc.

Los datos que se utilizaron para desarrollar esta aplicación fueron obtenidos de fuentes oficiales como la Dirección Nacional Electoral, el Ministerio del Interior, las plataformas políticas de cada lista electoral, el Ministerio de Salud, y datos del Censo Nacional de 2001 de Población y Vivienda (INDEC).

La aplicación, desarrollada en una plataforma abierta, fue generada el día antes de las elecciones de 2011, el 23 de octubre, con la participación de voluntarios de distintas especialidades.

Para la creación de este proyecto se trabajó con herramientas como Google Fusion Tables, Google Maps y bibliotecas de gráficos vectoriales. Se construyeron polígonos “para presentar mapeado geográfico y demografía electoral. Combinando polígonos en software GIS y geometrías de tablas públicas en las Tablas de Fusión Google”, se generaron tablas con “claves correspondientes a la base de datos electorales del Ministerio del Interior, Indra y datos socio-demográficos del INDEC”⁹². A partir de esa información se crearon visualizaciones con Google Maps.

⁹¹ «The Data Journalism Handbook.» 2011.

⁹² «The Data Journalism Handbook.» 2011.

El equipo de producción de esta aplicación publicó varios mapas temáticos que representaban la distribución espacial de la votación, representadas las zonas en tonos de color, donde la intensidad del color representaba el porcentaje de votos de las candidaturas presidenciales. Se hizo particular énfasis en los principales centros urbanos como la ciudad de Buenos Aires, los distritos del Gran Buenos Aires, Córdoba y Rosario.

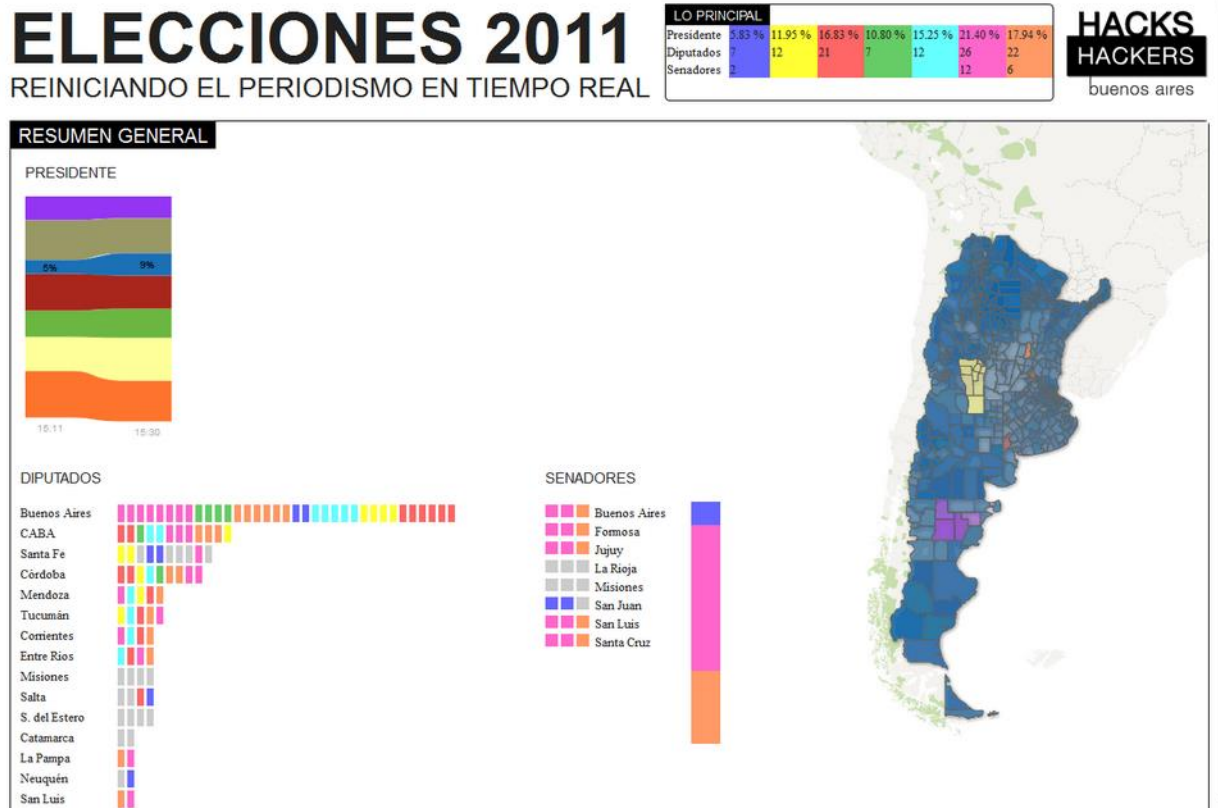


Ilustración 13 (Fuente: lanación.com) <http://www.lanacion.com.ar/1417158-una-manera-diferente-de-ver-los-resultados-electorales>

A su vez, también se realizaron mapas temáticos de elecciones anteriores, como las primarias presidenciales de 2011 y la elección presidencial de 2007. Se incluyeron también datos socio-demográficos (pobreza, mortalidad infantil y condiciones de vida) para realizar comparaciones.

“El proyecto también mostró la distribución espacial de las diferencias porcentuales de votos obtenidos por cada candidatura en la elección general de octubre, comparando con la elección primaria de agosto”⁹³. A continuación se creó un mapa animado que representaba el recuento de votos, “en el que se muestra el avance del mismo desde el cierre de la votación hasta la mañana siguiente”.

⁹³ «The Data Journalism Handbook.» 2011.

Este fue un proyecto que contaba con un espacio y tiempo específico, un período de vida reducido, pero que puede ser usado como base de datos o referencia para investigaciones futuras, ya sea de próximas elecciones o análisis de elecciones anteriores.

La plataforma *Hack Electoral* fue utilizada por los medios tradicionales, mediante coberturas televisivas, en la radio, en los medios impresos y online. Y sirvió de fuente y herramienta para posteriores análisis y noticias de las elecciones presidenciales de 2011 en Argentina. “El día de la elección, el sitio creado ese mismo día recibió alrededor de 20.000 visitantes diferentes y sus mapas fueron reproducidos en la tapa del diario *Página/12* dos días consecutivos, así como en artículos en *La Nación*. Algunos mapas aparecieron en las ediciones impresas del diario *Clarín*”⁹⁴. “*The Data Journalism Handbook*” dice que “fue la primera vez en la historia del periodismo argentino que se utilizó un despliegue interactivo de mapas en tiempo real”.

En los principales mapas publicados por *Hack Electoral* se podía ver claramente la abrumadora victoria de Cristina Fernández de Kirchner, la que ganó las elecciones de 2011 con un 54% de los votos. Esto se vio representado en la saturación del color que representaba a esta candidata.

La visualización en los mapas “también sirvió para ayudar a los usuarios a entender casos específicos donde candidatos locales tuvieron victorias por amplio margen en las provincias”⁹⁵.

Este caso específico de Periodismo de Datos demuestra que la fuerza de una visualización bien realizada y la potencia de los datos en crudo, puede generar un producto que sirva de referencia en el campo del periodismo, que sea repetido y propagado, y que le dé al lector una herramienta para entender en mayor profundidad el tema que se está analizando.

⁹⁴ «*The Data Journalism Handbook*.» 2011.

⁹⁵ «*The Data Journalism Handbook*.» 2011.

Capítulo V

PERIODISMO DE DATOS EN ROSARIO

En la ciudad de Rosario el Periodismo de Datos no es una disciplina muy desarrollada. Recién en los últimos años algunas entidades de la ciudad están empezando a utilizar los datos como una herramienta para generar noticias e investigaciones. Proyectos como la liberación de datos por parte de la Municipalidad de Rosario y el Gobierno de Santa Fe, se van abriendo paso para sentar las bases de esta disciplina en la ciudad.

Sin embargo -y a pesar- de las dificultades para conseguir datos y la poca cantidad de periodistas rosarinos interesados en el tema, existen dos entidades en la ciudad que trabajan actualmente con Periodismo de Datos. Estas son el diario digital **Porven** y la organización **Hacks/Hackers Rosario**.

En este capítulo del trabajo final voy a exponer ejemplos de proyectos de Periodismo de Datos realizados por estos medios, analizando los mismos con los conceptos planteados anteriormente en la tesis.

A fin de conocer en profundidad estas dos entidades y para poder llevar a cabo el análisis, he realizado entrevistas a **Horacio Bella** (director de Porven) y **Ezequiel Clerici** (co-fundador de Hacks/Hackers Rosario).

EL CASO DE PORVEN

Porven es un medio digital rosarino que se fundó a fines de 2014. Surgió como un “capricho” de sus creadores, los cuales veían que en la ciudad de Rosario los medios de comunicación hegemónicos no contaban con plataformas digitales bien desarrolladas o diseñadas, no eran atrapantes para los lectores y no aportaban nada nuevo. Por eso, **Horacio Bella** (su creador) junto con un equipo de desarrolladores, diseñadores y periodistas, fundó Porven.

Este medio nació digital y desde un primer momento dedicado al Periodismo de Datos. Horacio Bella afirma que “desde el día cero sabíamos que el Periodismo de Datos iba a ser una de las formas principales de cómo íbamos a mostrar las cosas, porque iba muy de la

mano con nuestra filosofía general, que es informar de una manera mucho más simple y con información fácil de consumir”⁹⁶.

Este medio tiene la singularidad de que su creador, Horacio Bella, no es periodista, sino desarrollador de productos digitales. Muchos de los que conforman hoy en día Porven no vienen de la disciplina periodística, sino que son formados en distintas áreas como la informática, diseño gráfico y programación; esto es lo que hace a este medio propicio para generar contenidos de Periodismo de Datos. En este trabajo anteriormente hablábamos de que una de las riquezas y fundamentos de esta disciplina es el trabajo en conjunto entre personas de distintas áreas, las que enriquecen y proveen conocimientos fundamentales para llevarla a cabo. Porven, en este sentido, es el único diario de Argentina que se fundó con la idea de hacer Periodismo de Datos y que, además, fue hecho por profesionales no periodistas.

Hemos hablado del caso de La Nación Data a nivel nacional, el cual es el referente de Periodismo de Datos en el país. Éste fue tomado por Porven para comenzar a idear proyectos de Periodismo de Datos. Horacio Bella dice: “cuando empezamos el proyecto, el diario La Nación no sé si ya estaba trabajando con el equipo consolidado de La Nación Data, pero sí estaba haciendo muchos proyectos de Periodismo de Datos, y fue su forma de trabajar uno de los referentes que nosotros tomamos a nivel nacional. A nivel internacional había otros como The Guardian y algunos medios de Estados Unidos”⁹⁷.

El fundador de Porven explica que “lo que principalmente hicimos, desde el primer momento en que propusimos el proyecto periodístico, fue empezar a relacionar cada una de las noticias con los datos, por más que no provengan de datos: cuantificarlos. Es decir, si una noticia sobre un accidente de tránsito venía desde las fuentes de forma no cuantificada, nosotros la estructurábamos” y agrega: “tratamos de cuantificar todas las cosas que sean cuantificables, por ejemplo: número de heridos, número de muertos, personas que iban en el vehículo en ese momento, etc. Entonces por más que la fuente de la cual nosotros obteníamos la información para generar la noticia no esté presentada en forma de datos, nosotros tratábamos de estructurarlos. Todo esto con el objetivo de en un futuro poder utilizar esa información que se va generando para crear nuevas noticias. Para comprobar, por ejemplo, por qué en cierto lugar de la ciudad se están generando tantos accidentes de

⁹⁶ Bella, Horacio, entrevista de Eugenia Michiels. (1 de Junio de 2017).

⁹⁷ Bella, Horacio, entrevista de Eugenia Michiels. (1 de Junio de 2017).

tránsito y si se debe a la cantidad de personas que van en ese auto, o la velocidad que se está yendo y demás”⁹⁸.

Horacio Bella asegura que ese fue el primer objetivo de Porven: tomar información no estructurada, convertirla a datos y reutilizarla para generar futuras investigaciones. Es decir: hacer Periodismo de Datos.

Hasta la fecha, Porven cuenta con dos proyectos de Periodismo de Datos publicados: un *Mapa de hechos* y un *Mapa de cortes programados*. A continuación vamos a analizarlos.

Mapa de hechos

Como su nombre lo dice, este mapa es un proyecto de Periodismo de Datos que muestra diferentes puntos de la ciudad de Rosario donde han ocurrido hechos específicos, los cuales se muestran geolocalizados y marcados en el mismo.

Por *hechos* Porven se refiere a sucesos ocurridos en la ciudad de Rosario que tienen relevancia pero que no necesariamente necesitan estar acompañados de una crónica o noticia. El diario hace una diferenciación entre un *hecho* y una *noticia*. Un hecho puede ser un accidente de tránsito, un robo, un asesinato, una manifestación, etc. es decir: todo lo que sea cuantificable. Algunos de estos hechos pueden estar acompañados por noticias o por crónicas, según la necesidad o importancia del mismo.

“El mapa funciona de base de datos y de registro general de todo lo que pasa. Nosotros consideramos que, por ejemplo, un accidente de tránsito por ahí no necesita una crónica o una noticia en sí, por eso solamente lo registramos como un hecho y lo dejamos guardado. Éste se puede compartir en las redes sociales, se puede generar información a partir de esa nota, la gente puede entrar a verla y demás; pero no consideramos que sea una noticia en sí que tenga relevancia dentro del sitio”⁹⁹ puntualiza Horacio Bella.

Esos hechos que aparecen marcados en el mapa conforman la base de datos propia del diario. Cada hecho es cargado en ese dataset para poder ser usado en próximas noticias o investigaciones. Además, los redactores de Porven cargan a la base de datos los hechos que

⁹⁸ Bella, Horacio, entrevista de Eugenia Michiels. (1 de Junio de 2017).

⁹⁹ Bella, Horacio, entrevista de Eugenia Michiels. (1 de Junio de 2017).

aparecen en las noticias que realizan día a día.

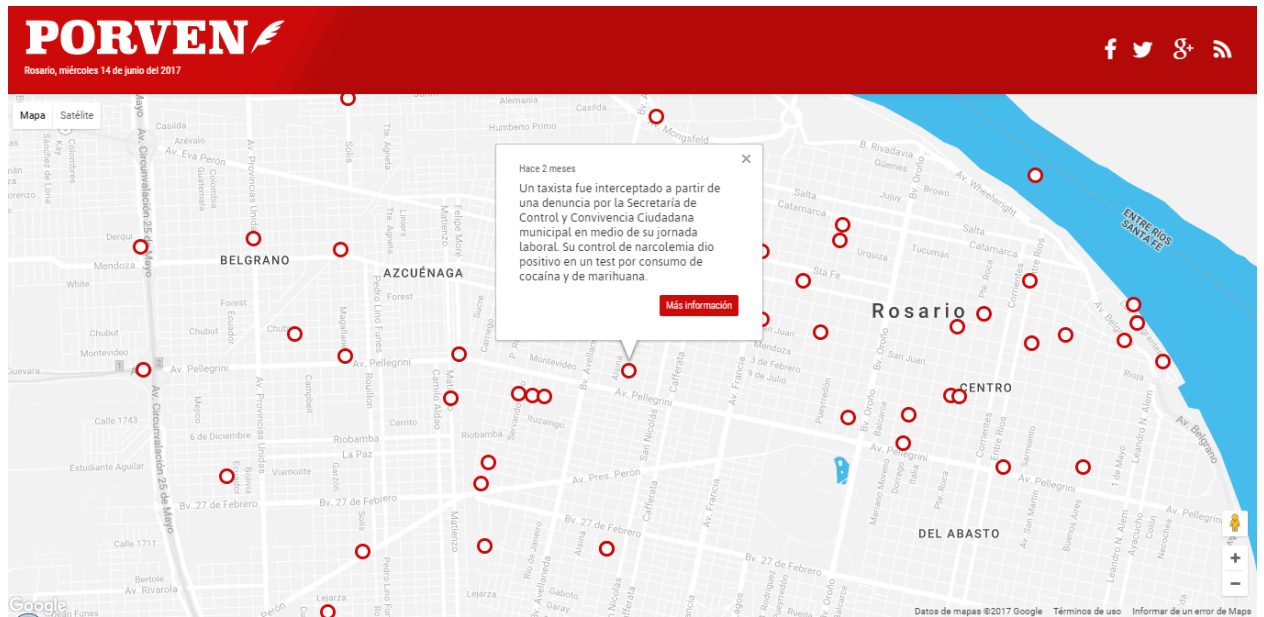


Ilustración 14 (Figura 1 - Fuente: Diario Porven)

En la **Figura 1** se puede observar un mapa de la ciudad de Rosario, visualizado a través de Google Maps. Los puntos rojos representan los hechos elegidos por Porven para ser visualizados; si se clickea en uno de esos puntos aparece un recuadro con una descripción del mismo, que incluye: momento en el que ocurrió (hace 2 meses, hace 2 horas, etc.), una breve descripción del suceso, y un botón de “más información” donde se puede acceder a una información más detallada del hecho.

Hace 7 meses

Hecho #28001

El Vicepresidente de **Newell's Old Boys** fue atacado a balazos desde un automóvil por un grupo de personas de identidad desconocida.

El hecho podría estar directamente relacionado con la interna de la barrabrava de Newell's.

Datos **3** delincuentes

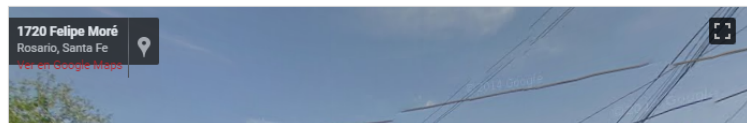


Ilustración 15 (Figura 2 – Fuente: Diario Porven)

En la **Figura 2** se observa lo que ocurre cuando se clickea en el botón de “más información”, el cual redirige al usuario a la página del diario en donde se encuentra una descripción más detalla del hecho. Allí se puede observar la fecha en que sucedió, el número del hecho, datos específicos como la cantidad de delincuentes involucrados o cantidad de heridos/muertes, la opción de compartir en las redes sociales, la posibilidad de ver la zona donde ocurrió el hecho a través de Google Street View, y una breve descripción del mismo.

Este **mapa de hechos** funciona no sólo como un servicio público para que el usuario pueda observar los hechos ocurridos en cada zona de la ciudad, sino también como una base de datos propia que puede ser utilizada para distintos proyectos a futuro. En este momento esos datos no están abiertos al público, es decir, si una persona quisiera tomar esos datos para realizar una investigación o visualización propia tendría que contabilizar los hechos uno por uno de forma manual. Es por esto que Horacio Bella asegura que la idea de Porven es abrir esos datos, generar un data set que pueda ser descargable para poder acceder a los datos en crudo, siguiendo la filosofía del Periodismo de Datos de hacer accesible al usuario las bases de datos utilizadas.

PROCESO DE CREACIÓN DEL MAPA

Como parte del proceso de creación de este *Mapa de hechos*, Porven llevó a cabo una serie de pasos que antes mencionábamos en este trabajo: recolección de los datos, limpieza de los mismos, análisis y visualización.

Respecto de la visualización, el equipo de Porven utilizó como base Wordpress, un gestor de contenidos, donde cargan las imágenes, datos y visualizaciones para darle forma al mapa. Para geolocalizar los hechos utilizaron un mapa de base donde marcaron cada suceso según sus coordenadas.

Como herramienta de visualización utilizaron Google Maps, conectando la base de datos cargada en Wordpress con el mapa. Esta es una visualización simple pero efectiva para lograr el resultado deseado.

Quiénes participaron del proyecto

Dijimos que Porven está conformado por programadores, diseñadores y periodistas, y que cada uno de ellos trabaja en conjunto para llevar a cabo los proyectos de Periodismo de Datos.

En este proyecto en específico los encargados de darle forma y sentido al mapa fueron: Lucas Pugliese (programador) y Horacio Bella, en la parte del desarrollo del panel en el cual se cargaron los datos; Sebastián Ferrari (diseñador) y Horacio Bella, en la parte del diseño y desarrollo del *front* y visualización del mapa; y el equipo de periodistas que trabajan en Porven, que se encargó de hacer un *data entry* de cada uno de los hechos visualizados en el mapa.

Herramientas utilizadas

- *Herramientas de recolección de datos:* Scraping de datos y lenguaje PHP.
- *Herramientas de filtro:* Wordpress y Maxwell.
- *Herramientas de visualización:* Google Maps y Google Street View.

Mapa de cortes programados

Otro de los proyectos de Periodismo de Datos realizado por Porven fue un **mapa de cortes programados**. Este mapa muestra las zonas en las que están programados cortes de luz por la empresa EPE (Empresa Provincial de la Energía de Santa Fe), visualizados fácilmente para que el usuario pueda encontrar su zona de residencia de manera más rápida, y de esa forma comprobar si va a ser afectado por el corte o no.

La particularidad de este mapa es que el usuario puede suscribirse con su correo electrónico para recibir información cuando el corte de luz esté programado en su zona de

residencia, de esta manera, Porven cumple la tarea de servicio público, brindándole al usuario una solución rápida para un problema diario.

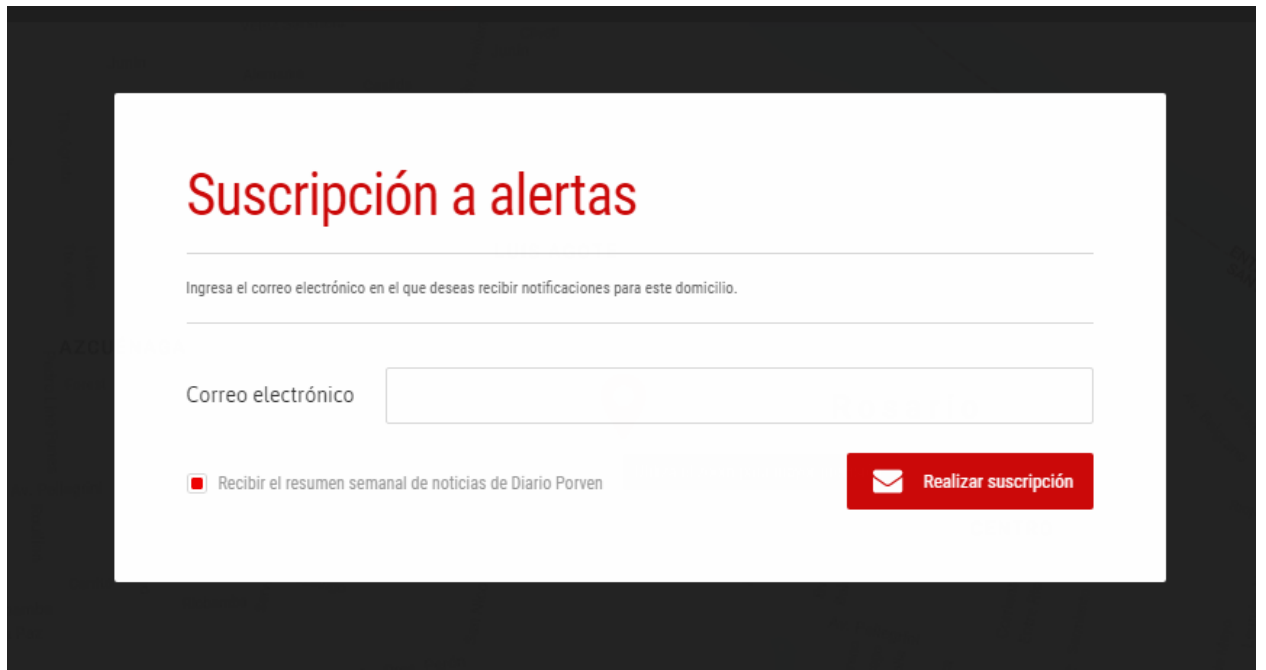
The image shows a web form titled "Suscripción a alertas" in red text. Below the title is a horizontal line, followed by the instruction "Ingresa el correo electrónico en el que deseas recibir notificaciones para este domicilio." Another horizontal line follows. Below this is the label "Correo electrónico" next to a white text input field. At the bottom left, there is a red square checkbox with the text "Recibir el resumen semanal de noticias de Diario Porven". At the bottom right, there is a red button with a white envelope icon and the text "Realizar suscripción".

Ilustración 16 (Figura 3 – Fuente: Porven)

En la **Figura 3** se puede ver cómo funciona el proceso de suscripción a las alertas de cortes programados. Se requiere un simple paso que es introducir un correo electrónico, e inmediatamente el usuario queda registrado en la base de datos de Porven para ser avisado de futuros cortes en su zona.

Lo que hizo Porven en esta investigación fue tomar los datos que proporciona la EPE – que son dados de manera no estructurada, desordenada y en forma de tabla o texto- y convertirlos a un formato más simple y entendible, visualizándolos en un mapa para que sean más fácil de comprender.

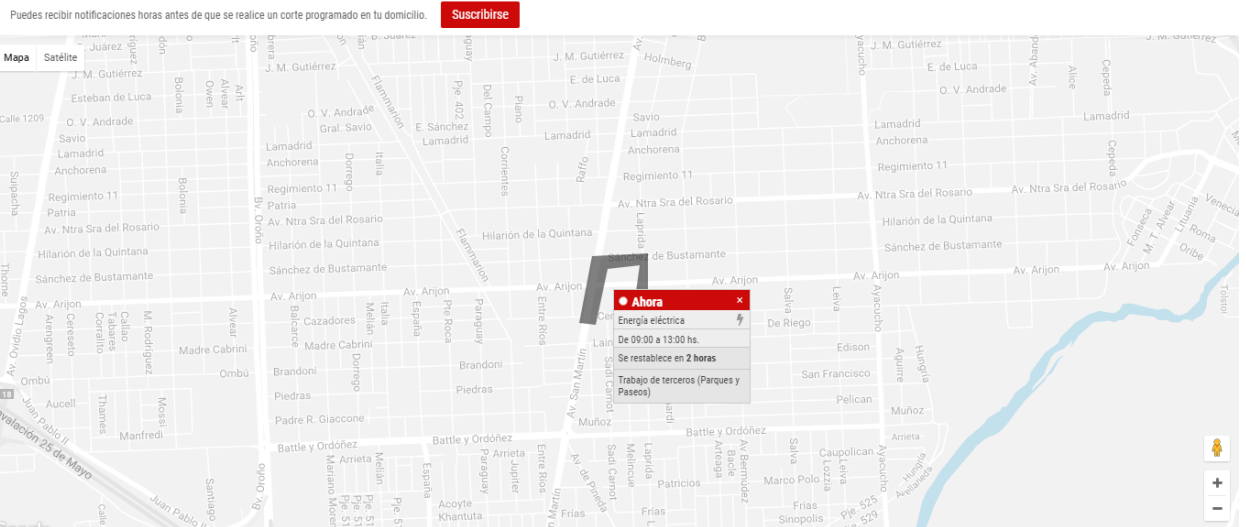


Ilustración 17 (Figura 4 – Fuente: Porven)

En la **Figura 4** se puede observar el Mapa de Cortes Programados, donde aparece visualizada mediante un sombreado el área que se encuentra afectada por un corte en el momento en que el usuario está accediendo a la información, en caso de que no haya ningún corte programado en ese momento, aparece visualizado un cartel que dice: “No hay cortes programados para hoy. Suscribite para recibir alertas de próximos cortes en tu domicilio”. En caso de que sí esté sucediendo un corte en alguna zona, en el mapa se especifica: el tipo de corte (energía eléctrica o agua potable), la franja horaria (ej: de 10:00 a 13:00hs), tiempo en el cual se restablecerá el servicio, y tipo de trabajo (ej: trabajo de terceros [parques y paseos]).

La EPE publica diariamente una lista de las zonas que van a ser afectadas por los cortes, el horario en que empieza el corte y el horario en que termina. “Nosotros consideramos que esa es información muy relevante, porque si yo sé que mañana se me va a cortar la luz, puedo tomar decisiones a partir de eso. Y al igual que nosotros determinamos que la información era muy relevante para las personas, también determinamos que la forma en la que se estaba entregando esa información no era la mejor. Por eso hicimos un sistema para ordenar esos datos: los organizamos de mejor manera, los guardamos en una base de datos y los mostramos de una forma mucho más rápida de entender” dice Horacio Bella, y

agrega: “Es básicamente la misma información que nos provee la EPE pero mostrada de una forma mucho más fácil de consumir”¹⁰⁰.

PROCESO DE CREACIÓN DEL MAPA

En este proyecto de Periodismo de Datos particularmente, Porven llevó a cabo una técnica de recolección de datos que mencionábamos anteriormente en este trabajo: la técnica del scraping.

Horacio Bella aclara que “generalmente hay una parte que no es scraping específicamente, pero está leyendo todo el tiempo esa información que publica la EPE y nos avisa cuándo hay un cambio, entonces nosotros podemos entrar ahí y cargar ya ese nuevo corte en nuestra base de datos” y agrega: “Hay partes que ya se cargan automáticamente y hay partes que nosotros tenemos que traducir”¹⁰¹. Es decir, si bien lo que hace Porven en este proyecto no es específicamente un scraping de datos, utiliza las técnicas del mismo para recibir información constantemente por parte del sistema de la EPE, la cual luego debe cargar en su base de datos.

Horacio Bella asegura que la parte más complicada de traducir esos datos es la forma en la que la EPE muestra las zonas afectadas por los cortes. Éstas son mostradas como una lista de calles en las que se nombran las intersecciones por zona. “Va armando toda una zona que vos tenés que formar en tu cabeza para saber si estás adentro o no. Eso es muy incómodo de leer, y para que cada persona no tenga que hacer eso lo hacemos nosotros con el mapa” explica el creador de Porven.

Esa es la tarea del periodista de datos: tomar un dataset no estructurado, convertirlo para que pueda ser utilizado y luego visualizarlo de manera que sea entendible.

Quiénes participaron del proyecto

Este proyecto fue llevado a cabo por el mismo equipo del *Mapa de hechos*, ya que todo el proceso de creación del mapa se hizo con una base de Wordpress, como ya dijimos. “La mayoría de los proyectos los hacemos sobre una base de Wordpress, porque eso ya nos da una base de datos bastante organizada, nos da la posibilidad de empezar a cargar datos en un panel que ya está armado. De esta manera, sobre esa base armamos un panel donde se genera el data entry. Hay partes del proceso que se hacen de manera automática a partir de

¹⁰⁰ Bella, Horacio, entrevista de Eugenia Michiels. (1 de Junio de 2017).

¹⁰¹ Bella, Horacio, entrevista de Eugenia Michiels. (1 de Junio de 2017).

los datos que publica la EPE, y hay partes que tenemos que organizar nosotros a mano”¹⁰² explica Horacio Bella.

Herramientas utilizadas

- *Herramientas de recolección de datos:* Scraping de datos y lenguaje PHP.
- *Herramientas de filtro:* Wordpress y Maxwell.
- *Herramientas de visualización:* Google Maps.

EI EQUIPO DE TRABAJO DE PORVEN

Uno de los aspectos fundamentales del Periodismo de Datos es el trabajo en equipo, entre programadores, diseñadores y periodistas. Puede suceder que un equipo de trabajo con estas características esté conformado por personas que no tienen los conocimientos básicos de Periodismo de Datos y las áreas que involucra, en este caso esa persona se va formando a medida que va entendiendo el funcionamiento y potencial de los datos.

Esto fue lo que ocurrió en el diario Porven, donde no sólo su creador no era un periodista de datos, sino que también su equipo de trabajo no tenía mucha idea de lo que era el Periodismo de Datos.

En Porven no existe un equipo específico dedicado exclusivamente al Periodismo de Datos como sí lo tiene el diario La Nación, sino que cada una de las personas que trabajan en el diario trata de aportar lo que puede para llevar a cabo los proyectos de Periodismo de Datos. Como mencionábamos anteriormente, el equipo encargado de realizar estos proyectos está conformado por Lucas Pugliese y Horacio Bella (programación), Sebastián Ferrari y Horacio Bella (diseño), y un grupo de periodistas encargados de la recolección de datos.

En este caso, los periodistas que conforman el equipo de trabajo tienen una formación en Periodismo de Datos que fue generada desde cero en el momento en que entraron a trabajar en el diario, formada a medida que avanzaba el desarrollo de los proyectos.

Porven se ha relacionado con Hacks/Hackers Rosario para trabajar en conjunto, utilizando así las habilidades periodísticas y dedicadas al Periodismo de Datos de las personas que conforman la organización.

Respecto del trabajo en conjunto con gente de distintas áreas, Horacio Bella recalca que es fácil trabajar de esta manera si se sigue cierta metodología, y agrega que a veces es

¹⁰² Bella, Horacio, entrevista de Eugenia Michiels. (1 de Junio de 2017).

difícil llegar a un acuerdo de cómo va a ser el producto final debido a los desacuerdos que se generan entre los participantes del equipo, respecto de la concepción e imaginación del proyecto en sí.

“El periodista puede tener una idea de cómo tiene que ser el producto, pero por cuestiones de diseño, por ejemplo, no se puede hacer. Lo mismo pasa con el programador, que realiza el producto de una forma que después al periodista no le termina sirviendo”¹⁰³ aclara el director de Porven. Es por esto que señala dos aspectos importantes a tener en cuenta:

1. En primer lugar los integrantes del equipo de trabajo deben conocer qué hacen los demás. “El periodista tiene que saber cuestiones básicas de diseño para no tener una idea que después no se pueda realizar, y el diseñador tiene que saber algo de periodismo para que la forma en que va a mostrar la información tenga sentido o vaya de la mano con la idea original del periodista”¹⁰⁴.
2. En segundo lugar, se introduce la figura del Project Manager, que es el encargado de supervisar el trabajo del equipo para que llegue a buen puerto. “El rol del Project Manager me parece muy importante. Tiene que ser una persona que conozca todas las áreas y que pueda hacer que el equipo funcione de manera ordenada y de forma dinámica”¹⁰⁵.

Porven trabaja de esta manera, con la figura del Project Manager que se encarga de llevar adelante el proyecto, y con un equipo de trabajo donde cada miembro conoce más o menos la actividad de sus compañeros.

“Si no sabés cómo funcionan las cosas es muy difícil que puedas resolver una visualización de un dato; es más, si no tenés los recursos para generar el producto digital, o la visión periodística para poder mostrarlo, capaz que estás delante de un montón de datos que son riquísimos pero no sabés utilizarlos”¹⁰⁶ agrega el director de Porven.

LAS BASES DE DATOS

Como Porven se pensó desde un primer momento como un medio que iba a utilizar el Periodismo de Datos como una de las principales herramientas para realizar noticias,

¹⁰³ Bella, Horacio, entrevista de Eugenia Michiels. (1 de Junio de 2017).

¹⁰⁴ Bella, Horacio, entrevista de Eugenia Michiels. (1 de Junio de 2017).

¹⁰⁵ Bella, Horacio, entrevista de Eugenia Michiels. (1 de Junio de 2017).

¹⁰⁶ Bella, Horacio, entrevista de Eugenia Michiels. (1 de Junio de 2017).

también desde un primer momento se comenzaron a cargar datos en un sistema de recolección y almacenamiento propio, es decir: se creó una base de datos.

Porven cuenta con una base de datos propia, donde se vienen cargando datos (principalmente los hechos que conforman su *Mapa de hechos*) a medida que se van realizando las noticias del día a día. En este momento el diario cuenta con dos años de datos cargados.

“No tenemos una base de datos fuera de lo que es la base de datos que utiliza el sitio, nosotros lo hacemos todo directamente dentro del sitio” dice Horacio Bella, y agrega: “La misma base de datos que nosotros utilizamos para que el sitio funcione la tenemos preparada para también poder editar y cargar los datos ahí”¹⁰⁷. Porven, en este sentido, encontró una solución práctica para un problema que es recurrente en el Periodismo de Datos: la falta de datos estructurados o la imposibilidad de encontrar los mismos.

Este diario desde un principio comenzó a cargar todos los datos que producía para así conformar una base de datos que no sólo nutre al sitio en sí y hace que funcione, sino que también proporciona material para ser recogido, analizado y visualizado.

“La base de datos que tenemos nosotros en el sitio nos permite, a través de un proceso de obtención de esos datos, poder filtrarlos” especifica el director de Porven, y aclara: “La base de datos está en formato Maxwell, que es un formato de base de datos que el sitio lee con PHP”¹⁰⁸.

Actualmente esta base de datos del diario no está disponible para ser descargada por el usuario, pero desde Porven tienen la idea de abrir ese dataset para que pueda ser descargable y editable por parte del usuario. A su vez también plantea la posibilidad de convertir sus mapas en mapas interactivos. “La idea es poder presentarlos de una forma que sea accesible y que cualquier persona los pueda descargar para refinarlos y utilizarlos”¹⁰⁹ aclara Horacio Bella.

LA CREDIBILIDAD DE LOS DATOS

Una de las características principales del Periodismo de Datos y una de las razones por las cuales es necesario comenzar a aplicarlo en todas las redacciones del mundo, es la credibilidad de los datos. A los datos no hay mucho que discutirle. Por eso las noticias e

¹⁰⁷ Bella, Horacio, entrevista de Eugenia Michiels. (1 de Junio de 2017).

¹⁰⁸ Bella, Horacio, entrevista de Eugenia Michiels. (1 de Junio de 2017).

¹⁰⁹ Bella, Horacio, entrevista de Eugenia Michiels. (1 de Junio de 2017).

investigaciones basadas en datos pueden aportar al periodismo un nivel de veracidad y credibilidad de la información que de otra manera no sería posible.

Proyectos como los llevados a cabo por Porven aportan a que la audiencia pueda acceder a una información comprobada y fiel a la verdad. Estos proyectos son necesarios, fundamentalmente ahora donde la credibilidad de las noticias está en tela de juicio por la enorme cantidad de información que provee la Web. “A través de las redes sociales se puede mostrar mucha información que nadie va a chequear; cosas que nunca se dijeron y se dan por válidas. Creo que el Periodismo de Datos tiene un poco de esa verdad o de esa dureza que tienen los números en sí; que la fuente sea un lugar fidedigno, que sea información que te entrega el Estado, información que te entregan las organizaciones”¹¹⁰ resalta Horacio Bella.

Respecto de las investigaciones a futuro que se propone realizar Porven en materia de Periodismo de Datos, su creador dice: “Creo que hay mucho trabajo para hacer, hay muchos datos disponibles. Creo que los productos que resultan de ese trabajo son noticias o investigaciones que a la gente le gusta consumir” y agrega: “Es un ámbito en el que el periodismo se tiene que respaldar mucho más, para solucionar un montón de problemas que está teniendo. A la gente hoy ya no le gusta leer una noticia de cinco párrafos donde se repite todo el tiempo lo mismo, con un solo dato que se presenta de forma repetitiva. Esas son cosas que puede resolver el Periodismo de Datos”¹¹¹.

EL PAPEL DEL PERIODISTA DE DATOS

Para entender un poco más la importancia del Periodismo de Datos y de su implementación en el periodismo del día a día, es preciso pensar el rol del periodista de datos como una especie de mutante que aporta mucho más que un periodista convencional. Es un mutante entre la tecnología y las viejas costumbres; mutante porque muta a un perfil periodístico mejorado y necesario en el tiempo actual.

“El trabajo del periodista de datos es traducir esa información que provee el Estado y las organizaciones –en forma de una tabla scv o un Excel- para que esos datos muestren lo que el usuario quiere saber, y lo muestre de una forma que sea fácil de consumir”¹¹² resalta Horacio Bella.

¹¹⁰ Bella, Horacio, entrevista de Eugenia Michiels. (1 de Junio de 2017).

¹¹¹ Bella, Horacio, entrevista de Eugenia Michiels. (1 de Junio de 2017).

¹¹² Bella, Horacio, entrevista de Eugenia Michiels. (1 de Junio de 2017).

Lo que dice el director de Porven es lo que analizamos anteriormente en este trabajo. Es necesario el trabajo del periodista de datos para traducir esa información que es dada a un formato que sea entendible por la audiencia. Sin el trabajo periodístico de por medio, esos datos no tienen ningún valor.

El Periodismo de Datos cambia el juego, genera un nuevo tipo de periodismo y un nuevo tipo de periodista. Este, como dice Horacio Bella, “deja de ser el periodista que está solo en su redacción encerrado en su oficina buscando, llamando por teléfono, averiguando historias o generando párrafos inmensos para publicar una columna. Cambia a este tipo de periodista nuevo que está sentado delante de un montón de bases de datos, de planillas de Excel, corroborando esa información con noticias que pasaron o con información que le hayan dado y que tiene que trabajar en conjunto con un montón de otras personas”¹¹³.

Al principio de esta tesis hemos tratado de definir qué es el Periodismo de Datos, propusimos una definición aproximada de lo que representa esta disciplina y analizamos sus diferentes características para poder alcanzar un nivel de comprensión más amplio. En esta etapa de nuestro análisis, es preciso saber qué definición puede dar un periodista que trabaja con datos sobre la disciplina que conforma las bases de Porven.

Horacio Bella define al Periodismo de Datos de la siguiente manera: “Es el trabajo de poder encontrar historias en los datos, que muchas veces están presentadas directamente en ellos y sólo hay que traducirlos. Es tomar esa cantidad de números que se encuentran en una base de datos y representarlos en un gráfico. Pero muchas veces también es poder encontrar historias en los datos, en eso que por ahí solamente es una barra en el gráfico, puede ser que detrás de esa información haya una historia”¹¹⁴.

EL CASO DE HACKS/HACKERS ROSARIO

El capítulo de Hacks/Hackers Rosario comenzó a funcionar en el año 2013, dos años después de que se inaugurara el capítulo de Buenos Aires, hasta ahora el más grande a nivel global y con la mayor cantidad de miembros activos.

Anteriormente mencionábamos que la organización Hacks and Hackers había nacido en Estados Unidos con el objetivo de generar un espacio de encuentro entre periodistas, programadores, diseñadores, desarrolladores web y todos aquellos interesados en repensar

¹¹³ Bella, Horacio, entrevista de Eugenia Michiels. (1 de Junio de 2017).

¹¹⁴ Bella, Horacio, entrevista de Eugenia Michiels. (1 de Junio de 2017).

al periodismo a través de herramientas y aplicaciones que se nutrieran de un trabajo interdisciplinario.

El capítulo de Rosario sigue esa misma lógica, centrándose principalmente en el desarrollo de proyectos periodísticos, ya que cada capítulo de Hacks/Hackers es autónomo y sigue fines acordes a sus miembros y sus ideas de proyectos, que pueden variar en su orientación: algunos se dedican más a emprender proyectos orientados al periodismo, otros persiguen un fin económico o emprendedor. En sí, Hacks/Hackers no es una organización sin fines de lucro, como describíamos anteriormente en este trabajo, sino que cada capítulo es autónomo y se cobija bajo el nombre de la organización para perseguir objetivos que pueden o no terminar produciendo ganancias.

Hacks/Hackers Rosario comenzó teniendo dos miembros cofundadores, Ezequiel Clerici (periodista rosarino que trabaja en Rosario3.com y en 5RTV) y Pablo Cuadrado (programador rosarino que trabaja actualmente en Quadra, una empresa de desarrollo de software y websites). El capítulo de Rosario fue fundado por estas dos personas, un programador y un periodista que trabajaron en conjunto para realizar el primer trabajo de Periodismo de Datos de Hacks/Hackers Rosario: un mapa de homicidios dolosos.

Hacks/Hackers Rosario surgió en un momento en que –como sucedió con el Periodismo de Datos- la información disponible en Internet no solo era cada vez más grande, sino que también comenzó a ser cada vez más accesible; comenzaron a surgir herramientas tecnológicas que abrieron puertas para que se desarrollaran nuevos espacios para realizar el periodismo y sobre todo nuevas formas de hacer periodismo, como lo es el Periodismo de Datos.

Hacks/Hackers Rosario nace en el año 2013 con una impronta periodística, fomentada por Ezequiel Clerici. Nace en una ciudad que presenta ciertas dificultades a la hora de hacer Periodismo de Datos, ya que el presupuesto es poco o nulo a la hora de invertir en este tipo de investigaciones, agregándole el factor de que los medios de comunicación de Rosario no cuentan con equipos de periodistas de datos o proyectos a futuro que involucren la implementación de esta disciplina –a excepción de Porven-, hecho que dificulta enormemente la posibilidad de desarrollar trabajos e investigaciones como los que tienen lugar en otros capítulos de Hacks/Hackers, como por ejemplo, el de Buenos Aires. Sin embargo, el capítulo de Rosario surgió a partir de una idea de proyecto, que se desarrolló y

tuvo mucho reconocimiento en la ciudad. Este fue el **Mapa de Homicidios Dolosos de Rosario**.

Mapa de homicidios dolosos

Este proyecto de Periodismo de Datos llevado a cabo por Hacks/Hackers Rosario fue publicado en el diario La Capital y trata, principalmente, de una investigación visualizada en un mapa donde se marcan todos los homicidios ocurridos en la ciudad de Rosario hasta la fecha, diferenciándolos por categorías según el tipo de arma o método con el que se cometió el asesinato, proporcionando a su vez información sobre la víctima.

El mapa fue dividido por Seccionales de la policía en las cuales se investigó cada caso, con las que se puede filtrar el contenido del mapa.

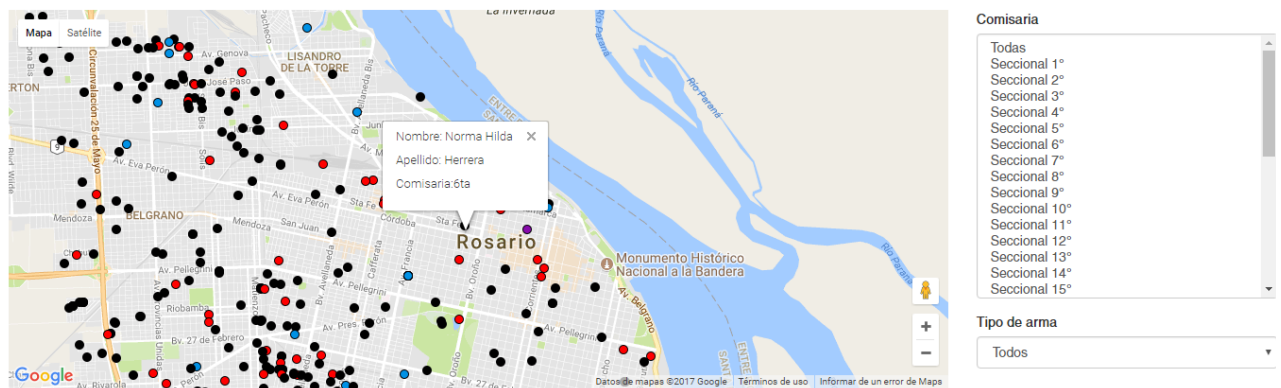


Ilustración 18 (Figura 5 – Fuente: <http://www.boasso24horas.com/index.php/mapa-de-homicidios-de-rosario>)

En la **Figura 5** se puede observar el mapa representado. Los homicidios aparecen marcados con puntos geolocalizados en el lugar del suceso, cada color representa un tipo de método de homicidio:

- **Negro:** Arma de fuego
- **Rojo:** Arma blanca
- **Celeste:** Traumatismo
- **Naranja:** Quemaduras
- **Violeta:** Ahorcamiento

Cuando se clickea en cada punto se puede ver un recuadro que proporciona detalles sobre la víctima: nombre, apellido y comisaría encargada del caso. El mapa se puede filtrar según el tipo de arma utilizada y según la seccional encargada del caso; de esta manera se divide por zonas los hechos, para observar más específicamente la cantidad y tipo de homicidios ocurridos en cada zona, como se puede observar en la **Figura 6**.



Ilustración 19 (Figura 6 – Fuente: <http://www.boasso24horas.com/index.php/mapa-de-homicidios-de-rosario>)

PROCESO DE CREACIÓN DEL MAPA

Este proyecto de Periodismo de Datos se llevó a cabo, en primer lugar, a partir de información proporcionada por el Ministerio de Seguridad de Rosario. Esta información les fue dada a Ezequiel Clerici y Pablo Cuadrado en forma de planilla, que contenía datos sobre todos los homicidios que se habían registrado en Rosario hasta el momento (año 2013): nombre y apellido de la víctima, tipo de arma utilizada, edad de la víctima, calle en donde sucedió el hecho, juzgado en donde se procesó el caso y comisaría a la que perteneció.

En base a esa información, el equipo de Hacks/Hackers se encargó de limpiar y filtrar ese dataset que estaba presentado en un formato no estructurado. “La policía o los ministerios tienen una forma de hacer ese tipo de Excel no pensando con una lógica de apertura de datos, sino como una forma de registro. Entonces nosotros le dimos esa estructura que necesitaba para ser leído sin ningún tipo de problema por un software de computadora”¹¹⁵ especifica Ezequiel Clerici.

Una vez convertidos esos datos a un formato legible por computadora, el equipo de Hacks/Hackers Rosario creó un software a medida para esta investigación, a través de una licencia GPL que es una licencia de software libre. A partir de ahí volcaron la información en un mapa de base con Google Maps, donde georeferenciaron cada uno de los homicidios, en base a las direcciones proporcionadas por el dataset con el que estaban trabajando.

Gran parte del trabajo realizado en este mapa fue hecho de forma manual, ya que para el momento en el que se llevó a cabo la investigación no sólo no existía la cantidad de herramientas de Periodismo de Datos que existen ahora, sino que también sus creadores, estos dos primeros miembros de Hacks/Hackers Rosario, Ezequiel Clerici y Pablo

¹¹⁵ Clerici, Ezequiel, entrevista de Eugenia Michiels. (21 de Junio de 2017).

Cuadrado, eran inexpertos en esta disciplina, por lo tanto no conocían todas las técnicas o atajos que podrían haberles facilitado el trabajo. “Ese laburo se hizo de forma manual en una etapa y de forma automatizada en otra, esto es porque era el primer trabajo a gran escala que hacíamos y, a medida que vas haciendo un primer proyecto, te vas dando cuenta de todos los errores que vas teniendo y los vas solucionando, los vas automatizando y los vas mejorando en base al error”¹¹⁶.

Una vez marcados y georeferenciados los homicidios, el equipo de trabajo se centró en mapear las comisarías asignadas a cada hecho. Lo que hicieron fue crear un polígono de la comisaría y de la zona que abarcaba, donde se pudo visualizar la cantidad de homicidios pertenecientes a cada zona. “A partir de ahí ya empezamos a saber que las comisarías que más crímenes tenían eran todas las que pertenecían a la herradura de la zona sur y el noroeste. En esa franja teníamos la mayor cantidad de crímenes”¹¹⁷. Gracias a la visualización de los datos se pudo diferenciar las zonas en donde ocurría la mayor cantidad de homicidios en la ciudad.

El siguiente paso de la investigación consistió en especificar en cada punto marcado en el mapa el nombre y apellido de la víctima, la intersección de la calle donde sucedió y el tipo de arma utilizada. A su vez, se buscaron los 214 crímenes ocurridos hasta la fecha en los registros de los diarios de la ciudad, donde el equipo de Hacks/Hackers Rosario encontró cerca del 80% de estos datos, la mayoría publicados en el diario La Capital. A partir de esa información, se comenzó a pensar en contactar a ese diario para publicar la investigación.

Repercusiones de la investigación

A partir de la visualización de los datos proporcionados por el Ministerio de Seguridad de Rosario, el equipo de Hacks/Hackers Rosario pudo demostrar, con el *Mapa de Homicidios Dolosos*, dos historias que surgieron a partir de esos datos.

La primera historia se basa en el descubrimiento de que existía, en el año 2013, un enorme problema de violencia gestado a partir de la venta ilegal de armas calibre 9mm, que son armas utilizadas por el ejército. Ezequiel Clerici explica esto: “Nosotros ya en el año 2013 sabíamos que Rosario había un problema gigantesco de armas, que la población podía acceder a armas de fuego calibre 9mm que es el calibre que usa la fuerza de seguridad o el ejército. Un civil no puede utilizar este tipo de arma. Estas eran armas reglamentarias que

¹¹⁶ Clerici, Ezequiel, entrevista de Eugenia Michiels. (21 de Junio de 2017).

¹¹⁷ Clerici, Ezequiel, entrevista de Eugenia Michiels. (21 de Junio de 2017).

usaba la fuerza de seguridad y el ejército”¹¹⁸. Con esto se puede deducir que alguien estaba vendiendo estas armas de manera ilegal en zonas de asentamientos irregulares; esto se pudo observar en el mapa cuando se comenzaron a filtrar los homicidios por zonas. “Era en barrios de marginalidad social donde sucedía la mayoría de los crímenes” asegura Ezequiel Clerici.

Esa fue la primera historia que reveló el *Mapa de Homicidios Dolosos*, la segunda se trataba de un conflicto social en las zonas donde ocurrieron los homicidios. Se reveló que la mayoría de esos crímenes eran cometidos por personas que se conocían, víctima y victimario, que tenían cierta familiaridad. Debido a la posesión cada vez mayor de armas de fuego por parte de las poblaciones marginadas de Rosario, se incrementó la violencia en estas zonas, donde los conflictos interpersonales de sus habitantes comenzaron a solucionarse de forma nefasta mediante el asesinato a mano armada. Esta revelación rompió con una mistificación y prejuicio que la ciudad de Rosario estaba atravesando para ese entonces, que era la afirmación, por parte de medios de comunicación -tanto locales como nacionales-, de que Rosario se había convertido en una ciudad insegura sólo por cuestiones de narcotráfico.

El *Mapa de Homicidios Dolosos* demostró que las razones de estos homicidios no estaban exclusivamente ligadas al narcotráfico. “La mayoría de esos crímenes no estaban atados a cuestiones de narcotráfico. Eso también era un dato muy importante, porque en 2013 había toda una estigmatización con el narcotráfico en la ciudad”¹¹⁹, y esta investigación logró romper con este mito.

Esta primera investigación de Periodismo de Datos llevada a cabo por Hacks/Hackers Rosario fue publicada en el diario La Capital, en forma de visualización interactiva con el *Mapa de Homicidios Dolosos*, y en forma de nota basada en datos, realizada por Hernán Lascano, que acompañó al mapa.

Quiénes participaron del proyecto

El proyecto estuvo a cargo de los dos cofundadores de Hacks/Hackers Rosario: Ezequiel Clerici y Pablo Cuadrado.

Ezequiel se encargó de conseguir el dataset del Ministerio de Seguridad en el que se basó la investigación, de filtrar los datos y estructurarlos.

¹¹⁸ Clerici, Ezequiel, entrevista de Eugenia Michiels. (21 de Junio de 2017).

¹¹⁹ Clerici, Ezequiel, entrevista de Eugenia Michiels. (21 de Junio de 2017).

Pablo estuvo encargado de la parte de programación y diseño del mapa. Ambos trabajaron en conjunto en el momento de geolocalizar los datos según su latitud y longitud de forma manual.

Herramientas utilizadas

- *Herramientas de recolección de datos:* Dataset proporcionado por el Ministerio de Seguridad de Rosario, no se utilizó una herramienta.
- *Herramientas de filtro:* Excel y Software libre (GPL).
- *Herramientas de visualización:* Google Maps.

EI EQUIPO DE TRABAJO DE HACKS/HACKERS ROSARIO

Hacks/Hackers, al no ser un medio de comunicación, está compuesto por miembros que vienen y van. Cualquiera puede participar de los hackatones o ser miembro del equipo, sin embargo, cuenta con un núcleo duro conformado por seis personas: Leticia Giaccaglia (comunicadora social y docente de la Universidad de Rosario), Ivana Mondelo (comunicadora social), Sebastián Sorribas (comunicador social y representante de Duck Duck Go en Rosario), Ezequiel Clerici (periodista), Rubén Acevedo (programador free lance) y Lucía Cuffia (periodista).

El capítulo de Rosario carece de programadores y diseñadores que trabajen de forma permanente en los proyectos. Hacks/Hackers Rosario cuenta con muchos periodistas y comunicadores interesados en el Periodismo de Datos, pero les resulta difícil atraer la atención de programadores y diseñadores, los cuales son fundamentales para trabajar proyectos basados en datos.

Otra característica del equipo de trabajo de esta organización es que en cada hackatón se forman grupos de trabajo, conformados usualmente por varios periodistas y al menos un programador y diseñador. Estos grupos realizan proyectos guiados por las propuestas que surgen en cada hackatón, donde se plantean las ideas y se comienza a trabajar. Los hackatones sirven como punto de conexión donde se forman estos equipos, pero luego el trabajo de cada uno de sus miembros es independiente a Hacks/Hackers Rosario. Ezequiel Clerici lo explica de esta manera: “Un hackatón es una jornada de trabajo intensa, se parte de un punto cero y se llega a una cantidad de serie de puntos que después se tienen que continuar y seguir por fuera de lo que es un hackatón; por lo tanto, ya queda dependiendo de ese equipo de trabajo. Inclusive Hacks/Hackers no tiene ninguna responsabilidad sobre

ese equipo, lo único que nosotros hacemos es generar el espacio y la propuesta para que venga gente a trabajar, después los equipos que se formen o la lógica que se arme ahí adentro es propia de cada uno de ellos”¹²⁰.

De esta manera Hacks/Hackers reafirma su posición como lugar de encuentro para que surjan proyectos de Periodismo de Datos.

Hacks/Hackers Rosario, al igual que el diario Porven, trabaja con la figura del Project Manager. En cada grupo de trabajo existe la figura del líder que, en este caso, cumpliría el papel de editor de proyecto. En este capítulo de Rosario el Project Manager suele ser un periodista, que establece determinadas pautas de comunicación, ésta figura de líder guía el proyecto, pero luego en la etapa en la que se plantea la forma en la que se van a visualizar los datos forman parte todos los miembros del grupo.

LAS BASES DE DATOS

Hacks/Hackers Rosario no cuenta con una base de datos propia. En el caso específico del *Mapa de Homicidios Dolosos* las bases de datos que se utilizaron fueron las brindadas por el Ministerio de Seguridad de Rosario, donde se creó una base de datos basada en esa información.

Sin embargo, esta organización tiene la idea de generar una base de datos propia. “Ese es un proyecto que estamos tratando de generar, esto de tener datasets ya curados y propios a disposición para que la gente pueda tomarlos y hacer proyectos; lo que pasa que eso implica tiempo y trabajo, y eso si no está remunerado tiene determinadas limitaciones, que es la limitación de laburar en tu tiempo libre”¹²¹ aclara Ezequiel Clerici.

Al no ser Hacks/Hackers un medio de comunicación, tiene limitaciones tales como el tiempo de trabajo del cual dispone cada miembro y la dedicación a los proyectos que puede proporcionar cada uno de ellos. Es por eso que en cada proyecto en específico se utilizan bases de datos externas.

Hacks/Hackers Rosario utiliza como base de datos Scraper, una herramienta para recolectar datos que mencionábamos anteriormente en este trabajo. También, Ezequiel Clerici asegura que “algunos programadores diseñan sus propios scrapers a medida, para buscar cierta cantidad de cosas, por ejemplo, para buscar tweets”¹²².

LA CREDIBILIDAD DE LOS DATOS

¹²⁰ Clerici, Ezequiel, entrevista de Eugenia Michiels. (21 de Junio de 2017).

¹²¹ Clerici, Ezequiel, entrevista de Eugenia Michiels. (21 de Junio de 2017).

¹²² Clerici, Ezequiel, entrevista de Eugenia Michiels. (21 de Junio de 2017).

Como decíamos con anterioridad, una de las principales características del Periodismo de Datos es que le da credibilidad a los datos y a las investigaciones basadas en ellos. Los proyectos llevados a cabo por organizaciones como Hacks/Hackers Rosario proporcionan información creíble y verificable.

Ezequiel Clerici recalca que esta característica del Periodismo de Datos es fundamental, dice que “Es una condición *sine qua non* del periodismo. Tenés que hablar con pruebas, tenés que hablar con hechos, tenés que hablar con datos que sean concisos, para darle credibilidad a tu artículo”.

El Periodismo de Datos no es nada nuevo, como venimos diciendo, pero surge en un momento en el que se dan ciertas condiciones que facilitan su trabajo. Una de ellas es la crisis de credibilidad por la que está pasando el periodismo, que, como dice Ezequiel Clerici, “tiene que ver con la masividad de los medios que hay, con las redes sociales en las que cualquiera dice cualquier cosa y se transforma en un interlocutor válido, con cómo se fue haciendo más líquida la credibilidad por las fake news, etc.”¹²³.

Otra de las razones por las cuales el periodismo ha perdido su credibilidad antaño indiscutible, es la masividad de información que se puede encontrar en Internet, aspecto que hemos analizado con anterioridad en este trabajo, y que reafirma Ezequiel Clerici cuando dice que “la lógica de internet está absolutamente masificada, hoy está inmersa en todo lo que se hace de manera productiva, y los medios de comunicación no escapan a esa lógica. Muchos de los periodistas que hoy toman la bandera del Periodismo de Datos son nativos digitales o milleniales, que se criaron en una concepción del mundo que es completamente distinta a la que tenían los viejos periodistas, del sistema analógico”¹²⁴. Esta lógica de internet sigue las premisas del Open Data, las de compartir el trabajo realizado o los procesos de trabajos, de abrir las fuentes, de trabajar en conjunto y de manera interdisciplinaria.

Lo que hace el Periodismo de Datos es brindarle herramientas al periodista para que pueda basar sus investigaciones en fuentes y datos confiables, para que pueda romper con esta crisis de credibilidad que atraviesa a todos los medios del mundo.

EL PAPEL DEL PERIODISTA DE DATOS

Después de todo lo que hemos analizado se llega a la conclusión de que no sólo ha cambiado la forma de hacer periodismo en un momento en donde la credibilidad del mismo

¹²³ Clerici, Ezequiel, entrevista de Eugenia Michiels. (21 de Junio de 2017).

¹²⁴ Clerici, Ezequiel, entrevista de Eugenia Michiels. (21 de Junio de 2017).

está en tela de juicio, sino que también ha cambiado el perfil del periodista. Ya no se trata de un periodista estático que se encuentra en la redacción de su diario esperando a recibir datos para llevar a cabo sus noticias e investigaciones, sino de un periodista que sale a buscar esos datos no a la calle sino en su computadora misma.

Como ha cambiado la lógica de hacer periodismo, también ha cambiado el rol que cumple un periodista. Los nuevos periodistas traen consigo un bagaje cultural arraigado en Internet, son nativos digitales y millenials que se han criado sabiendo manejar una computadora, como dice Ezequiel Clerici, son personas entre 20 y 40 años que siguen la lógica de internet, la de compartir la información. “El diferencial no está en tener los datos y ocultarlos, el diferencial es lo que haces con esa información, y tal vez si la abris vas a conseguir más gente que se sume a ese proyecto y que permita una mayor apertura para hacer algo mucho mejor”¹²⁵.

Respecto de la importancia de la implementación del Periodismo de Datos en los medios de comunicación y de la mutación del periodista hacia un periodista de datos, Ezequiel Clerici recalca una de las características que mencionamos varias veces en este trabajo: “Para mí una de las funciones primordiales del Periodismo de Datos es brindar un servicio público, le tenemos que dar a la gente algo que le sirva”¹²⁶.

Para terminar de entender qué es el Periodismo de Datos y qué función cumple en la creación de noticias e investigaciones, es preciso saber qué definición da Ezequiel Clerici sobre el mismo, la cual es la siguiente: “Es periodismo ayudado por computadora, que aprovecha las grandes cantidades de datos que tenemos disponibles a partir de la naturalización de Internet y a las filosofías de apertura de datos. Lo que se intenta hacer es aprovechar todo eso para generar historias que sigan teniendo como meta u objetivo el periodismo puro y duro, el periodismo tradicional; investigar en base a hechos reales, comprobables y verificables, y escribir a partir de eso un artículo que le permita al lector tomar una decisión que mejore su vida”¹²⁷.

Tanto Porven como Hacks/Hackers Rosario conforman el panorama de Periodismo de Datos en la ciudad de Rosario; uno publicando investigaciones de Periodismo de Datos en un medio de comunicación -como es el objetivo ideal de esta disciplina-, y otro generando

¹²⁵ Clerici, Ezequiel, entrevista de Eugenia Michiels. (21 de Junio de 2017).

¹²⁶ Clerici, Ezequiel, entrevista de Eugenia Michiels. (21 de Junio de 2017).

¹²⁷ Clerici, Ezequiel, entrevista de Eugenia Michiels. (21 de Junio de 2017).

espacios para que cada vez más gente de distintas disciplinas se acerque a esta especialización del periodismo que puede brindar tantos beneficios.

En la ciudad de Rosario, como dijeron Horacio Bella y Ezequiel Clerici, es muy complicado llevar a cabo un proyecto de Periodismo de Datos, ya sea por falta de interés de los medios de comunicación y las instituciones, por falta de liberación de los datos o por falta de presupuesto. A esto se le suma el factor de que muy pocos periodistas, programadores y diseñadores están interesados o interiorizados en esta disciplina.

Sin embargo, y a pesar de esto, estas dos instituciones siguen realizando proyectos de Periodismo de Datos y tratando de consolidar esta disciplina en la ciudad.

Conclusión

El objetivo de este trabajo fue el de poner en eje al Periodismo de Datos y analizar varios de sus aspectos, entre ellos: qué es, cómo se utiliza para generar noticias e investigaciones, qué herramientas usa para llevarlas a cabo, qué entidades significativas representan a esta disciplina y cuál es el panorama de la misma en Rosario. Para lograrlo se incluyeron ejemplos de investigaciones realizadas con Periodismo de Datos, más específicamente las llevadas a cabo por las entidades de esta disciplina pertenecientes a la ciudad de Rosario.

Luego de un análisis detallado de los proyectos y características de estas dos entidades rosarinas, teniendo en consideración la bibliografía consultada y las entrevistas realizadas para consolidar este trabajo, se llegó a formar varias conclusiones al respecto, las cuales he decidido dividir según las entidades analizadas.

Respecto del diario **Porven** se llega a la conclusión de que es el único medio de la ciudad de Rosario involucrado activamente en la utilización del Periodismo de Datos como herramienta para generar noticias e investigaciones. Su utilización de esta disciplina como una herramienta periodística se basa, principalmente, en la creación de mapas que cumplen la función de servicio público, un aspecto fundamental del Periodismo de Datos.

Una característica curiosa de Porven es que fue pensado como un medio periodístico dedicado a realizar investigaciones basadas en datos, hecho que no se puede ver en ningún otro medio del país. Este aspecto le concede el mérito de innovador, pero también lo pone en una posición riesgosa, en el sentido de que tiene que cumplir con ciertas expectativas. Si bien Porven es el único medio de Rosario dedicado activamente al Periodismo de Datos, los productos que realiza de esta disciplina todavía no están del todo desarrollados, ya que se basan únicamente en una forma de visualización: los mapas.

Para que Porven pueda consolidarse como un medio influyente de Periodismo de Datos a nivel regional y nacional, deberá no sólo realizar nuevas investigaciones de Periodismo de Datos sino también implementar otras formas de visualizarlas.

Por otro lado, **Hacks/Hackers Rosario** no es un medio de comunicación, sino una organización, pero cumple la función de mediador y proveedor de contenidos para llevar a cabo actividades e investigaciones relacionadas con el Periodismo de Datos. Con lo analizado se pudo concluir que el rol que cumple esta organización en la implementación

del Periodismo de Datos como herramienta para llevar a cabo noticias e investigaciones es fundamental, ya que es el punto de conexión entre personas y proyectos, es sumamente necesario para que esta disciplina se consolide en la ciudad de Rosario como una herramienta fundamental para llevar a cabo tareas periodísticas.

A nivel más general se pudo observar que se cumple la hipótesis de la investigación, la que afirma que el Periodismo de Datos cumple un papel importante en la producción de noticias a partir de la utilización de herramientas digitales y software. Ésta se cumple por dos cuestiones fundamentales, reveladas a partir de la entrevista realizada a Ezequiel Clerici.

La primera cuestión es en referencia a la crisis de credibilidad que sufren los medios a nivel global, la que se da por la enorme cantidad de información disponible en Internet y el mal manejo de la misma. En este trabajo se demostró –a través de la exposición de ejemplos de investigaciones de Periodismo de Datos y sus resultados, y a través de la explicación de qué herramientas utiliza esta disciplina y cómo las utiliza- que el Periodismo de Datos es una herramienta fundamental para terminar con esta crisis de credibilidad, ya que a los datos no se les puede cuestionar nada. Esta disciplina cumple el rol de verificadora de la información, y puede proporcionar noticias e investigaciones basadas en data comprobable y creíble.

La segunda cuestión es la utilización de grandes cantidades de datos para llevar a cabo esta disciplina. Si bien la irrupción de Internet como principal fuente de información para los periodistas significó que aumentara esta crisis de credibilidad de la información, también significó la implementación de nuevas herramientas digitales y software para poder manejar la gran cantidad de datos encontrados en la web. Uno de los objetivos de esta tesis fue el de explicar cómo el Periodismo de Datos utiliza las herramientas digitales y software para llevar a cabo sus investigaciones, y mediante la exposición de distintas herramientas a modo de ejemplo, se pudo concluir que éstas son un aspecto fundamental para realizar investigaciones basadas en datos.

El periodismo está sufriendo una crisis a nivel mundial, no sólo por estos aspectos que mencionamos, sino también porque la mayoría de los medios de comunicación todavía siguen una lógica que no se amolda a las características de la era digital. Con esta tesis se esclarece que se necesita implementar disciplinas como el Periodismo de Datos en todas las

redacciones del mundo, para alcanzar un nivel de información y de servicio acorde con las condiciones de la era digital, del Internet y del Open Data.

Respecto del panorama del Periodismo de Datos en la ciudad de Rosario, a la conclusión que se pudo llegar a partir de esta investigación es que todavía le falta mucho camino por recorrer. En Rosario la implementación del Periodismo de Datos como herramienta principal para generar noticias e investigaciones tiene muchas trabas, como hemos definido en el análisis de las entidades de la ciudad seleccionadas. Es menester empezar a pensar en estrategias comunicacionales en los medios de comunicación de Rosario que incluyan prácticas como el Periodismo de Datos, por muchas razones, pero por sobre todo para poder generar contenido único y confiable que se diferencie de Buenos Aires, el mayor productor de medios y contenidos periodísticos del país.

Para finalizar es preciso aclarar que esta tesis tuvo como objetivo definir las características del Periodismo de Datos y su aplicación en los medios de comunicación rosarinos, además de esto, ánimo de esta investigación también es haber logrado generar un acercamiento a los aspectos básicos de esta disciplina para que se comience a pensar en una forma de implementarla en los medios de comunicación de la ciudad, de ponerla en el eje de las discusiones sobre contenido de nuestros medios, y como una forma de iniciar una apertura hacia esta disciplina.

Bibliografía

- Alcántara, José F. *La sociedad de control. Privacidad, propiedad intelectual y el futuro de la libertad*. Barcelona: El Cobre Ediciones, 2008.
- Antón Bravo, Adolfo. «El periodismo de datos y la web semántica.» *CIC Cuadernos de Información y Comunicación*, 2013: 99-116.
- Baricco, Alessandro. *Los Bárbaros. Ensayos sobre la mutación*. Anagrama, 2006.
- Bella, Horacio, entrevista de Eugenia Michiels. (1 de Junio de 2017).
- Blanco Castilla, Elena, y Montse Quesada. *Periodismo de datos*. Tenerife: Cuadernos Artesanos de Comunicación, 2016.
- Bradshaw, Paul. *onlinejournalismblog.com*.
<https://onlinejournalismblog.com/2011/07/08/the-inverted-pyramid-of-data-journalism-in-spanish/> (último acceso: Abril de 2017).
- . «Scraping for Journalists.» Leanpub, 2013.
- Bueno Plaza, Ana. *Periodismo de datos. Antecedentes, ámbitos de desarrollo y aplicación en medios generalistas: el caso de El Confidencial y El Periódico*. Valladolid: Universidad de Valladolid, 2014.
- Cairo, Alberto. *El arte funcional. Infografía y visualización de información*. Madrid: Alamut, 2011.
- Cepci. *Un zettabyte de motivos para hablar de datos y desarrollo*. Bogotá D.C: Cepci, 2015.
- Chaparro Domínguez, María Ángeles. *La evolución del periodismo de precisión: el blog de The Guardian sobre periodismo de datos*. Logroño: Universidad Internacional de La Rioja, 2013.
- Clerici, Ezequiel, entrevista de Eugenia Michiels. (21 de Junio de 2017).
- Concha, Gastón ; Naser, Alejandra. *Datos abiertos: Un nuevo desafío para los gobiernos de la región*. Santiago de Chile, Santiago: Naciones Unidas, 2012.
- Crucianelli, Sandra. «¿Qué es el periodismo de datos?» *Cuadernos de periodistas*, 2013: 106-107-108-109-110-111-112-113-114-115-116-117-118-119-120-121-122-123-124.
- . *Herramientas Digitales para Periodistas*. 2013.
- Fallas, Hassel. «Tutorial: limpieza y análisis de datos.»
- Fernández García, Jorge Juan. *Más allá de Google*. Barcelona: Infonomía, 2008.
- Ferreras Rodríguez, E. M. «Aproximación teórica al perfil profesional del ‘Periodista de datos’.» 2013.
- Flores Vivar, Jesús Miguel. *Ecosistema del periodismo de datos. Cómo combinar fuentes, datos, contenidos, tecnología y mashups en el desarrollo del data journalism en la Red*. São Bernardo do Campo: C&S, 2012.
- Flores Vivar, Jesús, y Cecilia Salinas Aguilar. *El periodismo de datos como especialización de las organizaciones de noticias en Internet*. Madrid: Universidad Complutense de Madrid.
- Flores Vivar, Jesús; Salinas Aguilar, Cecilia. *El periodismo de datos como especialización de las organizaciones de noticias en Internet*. Universidad Complutense de Madrid, 2013.
- Foundation, Open Knowledge. *Manual de datos abiertos*. 2012.
- Hacks and Hackers. <https://hackshackers.com/> (último acceso: Mayo de 2017).
- Hidalgo, David ; Torres, Fabiola. «La navaja suiza del reportero.» Lima, Febrero de 2016.

Howard, Alexander Benjamin. *El arte y la ciencia del Periodismo de Datos*. Columbia Journalism School.

IBM. <https://www.ibm.com/developerworks/ssa/local/im/que-es-big-data/> (último acceso: Mayo de 2017).

Igarza, Roberto. *Burbujas de Ocio. Nuevas formas de consumo cultural*. Buenos Aires: La Crujía, 2009.

KnighFoundation. <https://knightfoundation.org/> (último acceso: Mayo de 2017).

La Nación Data. <http://data.lanacion.com.ar/> (último acceso: Mayo de 2017).

Lévy, Pierre. *Inteligencia colectiva*. Washington D.C: La Découverte (Essais), 2004.

Mancini, Pablo. *Hackear el periodismo*. Buenos Aires: La Crujía Ediciones, 2011.

Manovich, Lev. *What is Visualization*. 2010.

Mayer-Schönberger, Viktor, y Kenneth Cukier. «Big Data. La revolución de los datos masivos.» Madrid: Turner Publicaciones S.L, 2013.

Meyer, Philip. *The New Precision Journalism*. 1991.

Perry, Felipe; Paz, Miguel. *Manual de Periodismo de Datos Iberoamericano*. 2013.

Sabés Turmo, Fernando, y José Juan Verón Lassa. *Comunicación y la red. Nuevas formas de periodismo*. Zaragoza: Asociación de Periodistas de Aragón.

Salamanca, Nathalia, Jorge Luis Sierra, y Carlos Eduardo Huertas. *El periodista de investigación latinoamericano en la era digital*. ICFJ/Connectas, 2013.

Tapia, Alicia. *Big Data y periodismo de datos*. Madrid: Universidad Internacional de La Rioja.

«The Data Journalism Handbook.» 2011.

Wikileaks. (<https://wikileaks.org/>) (último acceso: Mayo de 2017).

Wikipedia. https://es.wikipedia.org/wiki/Big_data (último acceso: Mayo de 2017).

Word Cloud. <http://www.wordclouds.com/> (último acceso: Mayo de 2017).

Word Tree. <https://www.jasondavies.com/wordtree/> (último acceso: Mayo de 2017).

Zanoni, Leandro. *El imperio digital. El nuevo paradigma de la comunicación 2.0*. Buenos Aires: Ediciones B, 2008.

Anexo

ENTREVISTAS PERSONALES

Entrevista a Horacio Bella – Diario Porven

1 de Junio de 2017

Entrevistador: Eugenia Michiels

Lugar: Rosario, Santa Fe

1. ¿Cómo utiliza Porven los datos para generar noticias?

Nosotros lo que principalmente hicimos desde el primer momento que propusimos el proyecto periodístico fue empezar a relacionar cada una de las noticias, por más que sea algo realmente que no provenga de datos, llevar esos datos a la cuantificación, de cuantificar obviamente todas las cosas que sean cuantificables: número de heridos, número de muertos, personas que iban en el vehículo en ese momento; o en una noticia sobre un robo: la cantidad de dinero que han robado, la cantidad de delincuentes que había, cantidad de armas que se usaron y demás. Entonces, por más que la fuente de la cual nosotros obteníamos la información para generar la noticia no esté presentada en forma de datos, nosotros tratábamos de convertirlos. Todo esto con el objetivo de en un futuro poder utilizar toda esa información que se va generando para generar nuevas noticias; esto nos permitía por ejemplo, que con una noticia, o una serie de noticias de accidentes de tránsito en cierta zona de la ciudad, nosotros después podíamos cuantificar esas noticias y tomar esas noticias en particular para poder generar una nueva información. Por ejemplo, por qué en cierto lugar de la ciudad se están generando tantos accidentes de tránsito y si se debe a la cantidad de personas que van en ese auto o la velocidad que se está yendo, y demás.

Ese fue el primero objetivo. Por más que la información que nosotros obteníamos no estaba en forma de datos, tratamos de convertirlos para en un futuro poder generar nuevas noticias a partir de esos datos que nosotros teníamos.

2. ¿Y esos datos con qué herramientas los convirtieron?

Hay varias formas, nosotros tratamos de manejar todo dentro del sitio de una base de datos. La forma en que planteamos la base de datos que tenemos nosotros en el sitio nos

permite poder filtrarlos a través de un proceso de obtención de esos datos. Lo podemos obtener solamente a través de un filtro que nosotros tenemos, a partir de ahí filtrar solamente las noticias de accidentes de tránsito donde hubo muertos. Esa información está en formato Maxwell, que es un formato de base de datos; el sitio lo lee con PHP, que es una tecnología que te devuelve el formato para mostrarlo en la web. Básicamente lo que es el sitio en sí nosotros nos manejamos con PHP para leerlo y Maxwell para guardar la información en la base de datos.

3. Y eso lo visualizaron en el Mapa de Hechos que hay en el diario.

Exacto. Nosotros paralelamente en el sitio mostramos un mapa de lo que llamamos “hechos”, todas estas cosas que son cuantificables: accidentes de tránsito, robos, asesinatos, manifestaciones, todo lo que podamos medir los datos y principalmente geolocalizarlos. Esa es otra cosa que hacemos también con las noticias, todas las noticias que se puedan geolocalizar, entonces así tenemos varias formas de acceder a esa información. Nosotros la mostramos en el mapa con el punto en donde se generó esa noticia.

4. Me llamó la atención que cada hecho aparece descripto brevemente pero no hay un link a la noticia que se relaciona, ¿o no tienen relación con una noticia específica?

Hay una diferenciación que nosotros hacemos, hay algunas noticias que sí tienen que tener su crónica o su desarrollo y hay algunos hechos que a nosotros nos interesa cargarlos en el sitio más que nada para tener una referencia y después medirlo de forma más precisa, pero que no tiene su crónica. Separamos lo que es una noticia común de lo que nosotros llamamos un “hecho”. Un hecho puede ser un accidente, un robo, un asesinato, una manifestación, etc.

5. El Mapa de Hechos sería una base de datos para Porven y para la gente.

Exacto, el mapa funciona de base de datos y de registro general de todo lo que pasa; pero nosotros consideramos que, por ejemplo, un accidente de tránsito capaz que no necesita una crónica o una noticia en sí, solamente lo registramos como un hecho y lo dejamos guardado. Se puede compartir el hecho en las redes sociales, se puede generar información a partir de esa nota, la gente puede entrar a verla y demás, pero no consideramos que sea una noticia en sí como para que tenga relevancia dentro del sitio, entonces, a partir de ahí, nosotros generamos una división entre lo que es un hecho y lo que es un hecho que merece ser noticia.

6. Sé que usaron Google Maps para generar el mapa ¿qué otras herramientas utilizaron para ubicar cada hecho en su lugar o para formar esa base de datos?

Nosotros tenemos todo y tratamos de generar todo dentro del mismo panel. Usamos una especie de base de Wordpress, que es un gestor de contenidos, para todo lo que sea carga de imágenes, carga de información y demás; y adaptamos a eso toda una base de datos en la cual podemos cargar los datos y geolocalizar la información. Para geolocalizarlos usamos un mapa de base que es muy fácil cargarlo desde el panel: arrastrás un marcador hasta el lugar en donde pasó el hecho y ya se guardan las coordenadas, se guarda un dato de longitud y otro de latitud, y eso después te da un punto específico en el mapa. Después para mostrarlo sí utilizamos Google Maps, conectamos ese Google Maps con la base de datos que nosotros tenemos guardada.

7. ¿Quiénes participaron de este proyecto en específico?

De la parte del desarrollo del esquema del panel en el cual nosotros cargamos la información participó Lucas Pugliese, que es un programador, y Horacio Bella. En la parte de diseño también estuve trabajando yo, y estuvo trabajando Sebastián Ferrari en algunos detalles de diseño y el desarrollo de lo que se llama *front*, que es básicamente lo que se muestra en la pantalla de la computadora y los celulares. En la parte de datos trabajaron los chicos que son los redactores de Porven, que con cada noticia ellos van cargando cada uno de los datos, hacen un *data entry* de cada uno de los hechos que pasan.

8. Sé que ustedes también tienen otro proyecto que es el Mapa Cortes Programados, ¿cómo es ese proyecto?

Nosotros en ese proyecto en particular tomamos los datos que provee la EPE sobre los cortes programados. Ellos publican diariamente una lista de la zona que va a estar afectada por el corte, el horario que empieza y el horario en que termina el corte programado. Nosotros consideramos que esa es información muy relevante, porque creo que si yo sé que mañana se me va a cortar la luz a la noche en mi casa, puedo tomar decisiones a partir de eso. Y al igual que nosotros determinamos que la información era muy relevante para las personas en sí, también determinamos que la forma en la que se estaba entregando esa información no era la mejor, entonces hicimos un sistemita que toma esos datos y los ordena un poco mejor, porque la EPE los entrega de una forma muy desprolija. Hay una parte que se hace automática y una parte que la hacemos de forma manual; los organizamos

de mejor manera, los guardamos en una base de datos y los mostramos de una forma mucho más rápida de entender que es sobre un mapa, donde te marca cuál va a ser la zona afectada, el horario en el que va a pasar el corte y el horario en que va a terminar. Es básicamente la misma información que nos provee la EPE pero mostrada de una forma mucho más fácil de consumir, y además a eso le agregamos la opción de que te pueda llegar por correo electrónico un aviso un día antes de que se corte la luz. Eso fue básicamente lo que hicimos, tomamos datos que estaban muy desorganizados, los procesamos y los mostramos de una forma mucho más fácil de consumir.

9. ¿Quién participó de este proyecto?

Básicamente las mismas personas que el anterior, también lo hicimos sobre una base de Wordpress. La mayoría de los proyectos los hacemos sobre una base de Wordpress, porque eso ya nos da una base de datos bastante organizada, nos da la posibilidad de empezar a cargar datos de un panel que ya está armado; así que sobre esa base de Wordpress armamos un panel donde se genera el *data entry*, hay partes que se hacen automáticas desde esos datos que publica la EPE y hay partes que tenemos que organizar nosotros a mano, y después se publica, así que las personas que trabajaron fuimos los mismos.

10. ¿Utilizan la técnica del scraping de datos?

En este caso de los cortes programados sí, generalmente hay una parte que no es scraping particularmente, pero que está leyendo todo el tiempo esa información que publica la EPE y nos avisa cuándo hay un cambio. Nosotros podemos entrar ahí y cargar ya ese nuevo corte. Hay partes que ya se cargan automáticamente y hay partes que nosotros tenemos que traducir. En realidad, la parte más complicada de traducir es la forma en la que se muestra la zona que va a estar afectada, porque eso hoy la EPE lo está mostrando con una lista de calles en las que te va dando la idea de intersecciones, por zona. Va armando toda una zona en la que vos tenés que andar formándola en tu cabeza para saber si estás adentro. Eso es muy incómodo de leer, y para que cada persona no tenga que hacer eso lo hacemos nosotros una vez, lo dibujamos en el mapa y eso ya se carga, entonces vos después lo ves como una zona geográfica nada más.

11. ¿Qué bases de datos utilizan para realizar estos proyectos?

Lo que hacemos con los cortes programados en particular es tomar eso que ni siquiera es una base de datos. La forma en que se presenta es muy difícil de leer, está presentado de una forma muy desprolija, prácticamente es un párrafo pasado a tablas y a nosotros no nos

sirve para guardar la información. Eso lo guardamos nosotros en una base de datos Maxwell. No tenemos una base de datos fuera de lo que es la base de datos que utiliza el sitio, se podría hacer obviamente todo el trabajo de generar esa base de datos y procesarla, refinarla antes de agarrar la información, pero nosotros lo hacemos todo directamente dentro del sitio. La misma base de datos que nosotros utilizamos para que el sitio funcione la tenemos preparada para también poder editar y cargar los datos ahí.

12. ¿Planean en algún momento liberar esa base de datos que tienen propia o armar una base de datos bien prolija?

Sí, nosotros lo que tenemos cargado es dos años de todos los hechos que pasaron en Rosario, todo lo que te comentaba antes. Estamos cargando inclusive cosas que son noticias y cosas que consideramos que son hechos pero que no son noticias, los tenemos cargado, y hoy en día está disponible todo en el mapa. Lo que sí es difícil acceder para procesarlo, nosotros obviamente tenemos la idea de que eso esté disponible para que una persona lo descargue y lo pueda editar, estamos trabajando en eso y queremos, además, que se pueda descargar para que cualquier persona con Periodismo de Datos lo pueda editar, y que además se pueda interactuar en el mapa.

13. ¿Eso en este momento no existe?

En este momento no se puede, solamente en este momento en el mapa se pueden ver todos los hechos que sucedieron y que tenemos cargados, pero sin interacción. No podés filtrar, no podés buscar un hecho en particular. La idea es poder presentarlo de una forma que sea accesible y que cualquier persona lo pueda descargar para refinarlo y utilizarlo.

14. ¿Porven cuenta con un grupo formado de Periodismo de Datos?

Los chicos que están trabajando en la parte periodística tienen su formación que la hemos generado desde cero nosotros, no han sido personas que han llegado con una formación de Periodismo de Datos, sino que los hemos formado en todo el proceso de desarrollo del proyecto. Sí hemos participado en algunos proyecto en conjunto con la gente de Hacks/Hackers Rosario que sí tienen por ahí esa formación, y la idea es, como comentabas antes, con todos los datos que está publicando la Municipalidad y que el Gobierno de Santa Fe va a empezar a publicar ahora, poder utilizar esos datos externos para poder generar nuevas noticias o para al menos poder procesarlos y mostrarlos de una forma que también es parte por ahí de la filosofía que tenemos nosotros: poder tomar cosas que

son difíciles de leer y hacerlo un poco más accesible para que el consumidor final pueda nutrirse de esa información de una manera simple.

15. Lo que veía es que ustedes tienen una visualización más que nada centrada en mapas, todavía no se centran en lo que es por ahí una infografía interactiva.

Sí, la mayoría de los proyectos que estuvimos trabajando fueron relacionados a mapas, tratamos de mostrar la información en mapas, particularmente porque justo los proyectos que nos habían interesado daban para ese perfil, no por nada en particular.

16. ¿Planean hacer otro tipo de visualizaciones que no sean mapas?

Sí, la verdad es que, como la base de Porven es muy tecnológica, es más que nada orientada al diseño y al desarrollo, ese fue un pilar bastante importante al margen del pilar periodístico. Eso es lo que más nos permite generar el medio de una forma distinta a los demás. Es una idea que teníamos siempre pendiente pero siempre presente, para mostrar la información, porque nosotros consideramos que principalmente el Periodismo de Datos te permite mostrar la información de una manera más clara, mostrar algunas cosas que realmente necesitan un párrafo pero muchas cosas es más fácil explicarlas a través de datos, a través de imágenes, gráficos, infografías y mapas. De esas cosas nosotros nos nutrimos para cumplir el objetivo que por ahí tiene particularmente nuestro medio, que es informar de una manera mucho más simple.

17. ¿Tenían pensado desde un principio hacer Periodismo de Datos?

Fue desde un principio y por eso lo planteamos de esa forma de empezar desde el día cero a guardar toda la información de una manera que en un futuro la podíamos utilizar, así que ese fue uno de los pilares a la hora de crear el medio, saber que guardar los datos de esa forma nos iba a permitir en un futuro poder utilizarlos.

18. ¿Cómo trabajan los diseñadores, programadores y periodistas en conjunto dentro de Porven?

Nosotros trabajamos, al margen de la parte de periodismo, mucho en la parte, yo particularmente, de desarrollo web. Y en todo ámbito donde hay tantos aspectos para trabajar, diseño, parte periodística, parte de desarrollo, parte de desarrollo también del *front* y del *back*. *Front* nosotros le llamamos a lo que vos realmente ves en el navegador, con lo que interactúa el usuario final. Y el *back* es donde se cargan los datos, cómo funciona la base de datos, cómo funciona el desarrollo de la programación y demás. Y cada una de esas

áreas está compuesta por al menos una o dos personas, entonces es muy difícil que trabajen en conjunto. En realidad, es fácil que trabajen en conjunto si se sigue una cierta metodología, pero es muy difícil saber y poder llegar a un acuerdo sobre cómo va a ser el producto final sabiendo qué puede ofrecer cada uno, porque el periodista puede tener una idea de cómo tiene que ser el producto pero por cuestiones de *front* no se puede hacer o por cuestiones de *back* no se puede hacer, por cuestiones de diseño no se puede hacer. Y lo mismo con el programador, que lo hace de una forma que después al periodista no le termina sirviendo. Hay dos cosas muy importantes en esto; lo primero se relaciona con toda esta evolución, tanto del periodismo como del trabajo de desarrollo de productos digitales, que es que cada uno de los integrantes del equipo conozca qué hacen los demás. El periodista tiene que saber cuestiones básicas de diseño para no tener una idea que después no se pueda realizar, el periodista tiene que saber cuestiones básicas de *back* para que no piense algo que no se pueda realizar, y la persona que hace el *front* también tiene que saber diseño, y el diseñador tiene que saber algo de periodismo para que la forma en que vaya a resolver cómo se va a mostrar esa información realmente tenga o vaya de la mano con la idea que tuvo el periodista. Esa parte en la que todo el equipo, además de trabajar en conjunto, sepa qué puede hacer el otro y cómo funciona el trabajo del otro me parece muy importante en este tipo de equipos de trabajo. Y segundo, el puesto de Project Manager, que es un puesto que funciona en varias áreas, principalmente en la parte de desarrollo de productos digitales como es el tipo de proyectos de Periodismo de Datos. Básicamente es el desarrollo de un producto digital con la inserción del periodismo en sí. Este rol de Project Manager me parece muy importante, tiene que ser una persona que conozca de todos los puntos y que pueda hacer que este equipo funcione de forma ordenada y de forma dinámica. Nosotros siempre trabajamos de esa manera e intentamos que eso pase, que haya una persona que sea el Project Manager, la que lleve adelante el proyecto y que tenga por ahí una jerarquía un poco superior, al menos en ese proyecto en particular, para que pueda tomar decisiones que hagan que el producto salga, porque muchas veces pasa eso, entre la idea del que lo propuso y el que no lo pudo hacer hay un cortocircuito que genera que el producto no salga o no salga de la manera en la que finalmente se propuso.

Si vos conocés qué es lo que se puede hacer, cuál es la mejor forma de mostrarlo y qué es lo que finalmente va a ser un producto bueno para consumir, vos desde el vamos podés tirar la idea. Si vos no sabés cómo funcionan las cosas es muy difícil que puedas resolver

una visualización de un dato, es más, si vos no tenés los recursos para generar el producto digital o la visión periodística para poder mostrarlo, capaz que estás delante de un montón de datos que son riquísimos pero no sabés utilizarlos, no se te ocurre cómo eso puede presentarse de una forma que sea rica de consumir.

19. ¿La idea de empezar a utilizar noticias basadas en datos surgió cuando surgió el diario? ¿Creés que el Periodismo de Datos es una herramienta útil para generar nuevos proyectos y mejores investigaciones?

Yo creo que sí, particularmente por varios factores, creo que hoy en día donde la credibilidad de las noticias está tan en tela de juicio, donde es tan fácil presentar información falsa como verdadera o como verosímil al menos y que eso se esparza de una manera viral casi. Leí un artículo, del New York Times creo que era, que decía que más de la mitad de las noticias que se consumieron durante la campaña presidencial en Estados Unidos eran falsas. A través de las redes sociales se puede mostrar mucha información que nadie va a chequear, nadie va a corroborar, cosas que nunca se dijeron y se dan por válidas, cosas que nunca pasaron que se dan por válidas; creo que el Periodismo de Datos tiene un poco de esa verdad o de esa dureza que tienen los números en sí, digamos, que la fuente sea un lugar fidedigno, que sea información que te entrega el Estado, información que te entregan las organizaciones. Y al dato no hay mucho que discutirle, es ese número que obviamente interpretado de una forma u otra puede cambiar el significado que tenga, pero no se le puede discutir la verosimilitud. Creo que el Periodismo de Datos lo que hace es analizar esa cantidad enorme de datos. Hay mucho trabajo para hacer ahí, hay muchos datos disponibles. Hay mucho trabajo por hacer y creo que los productos que resultan de ese trabajo son notas, o son noticias, o son investigaciones que a la gente le gusta consumir, son fáciles de consumir y no tienen ese problema de la poca credibilidad; eso es lo que me parece que es un ámbito en el que el periodismo se tiene que respaldar mucho más para solucionar un montón de problemas que está teniendo. A la gente hoy ya no le gusta leer una noticia de cinco párrafos donde se repite todo el tiempo lo mismo y hay un solo dato que se presenta de forma repetitiva en un montón de párrafos de forma aburrida, creo que hay que darle una vuelta a cómo se presenta esa información. Y después está el otro tema como te decía antes de la credibilidad. Así que esas son cosas que puede resolver el Periodismo de Datos.

20. Incluso la persona que vea esa base de datos puede generar su propia historia con esos datos.

Hay un trabajo en el medio que tiene que hacer el periodista, me refiero como periodista a este grupo de trabajo del que hablábamos, entre esa información que provee el Estado o la organización en particular que para el consumidor final es un montón de números tirados en una tabla en un scv o un Excel, que no le significan nada. El trabajo del periodista en el medio es traducirle esa información, no solamente traducirlo sino hacerlo de una forma que realmente muestre lo que la persona quiere saber y lo muestre de una forma que sea fácil de consumir, me parece que ese es el trabajo para hacer, porque sin el trabajo del periodista de por medio los datos que provee el Estado no son nada. Si en todos los datos que publica la Municipalidad o los que va a publicar el Gobierno de Santa Fe no hay un trabajo periodístico en el medio no tienen ningún valor. Un usuario común no va a agarrar una base de datos y se va a poner a revisar los presupuestos. Creo que el trabajo de periodismo que hay en el medio es lo que realmente le da valor a esos datos.

21. ¿Cuál sería tu definición de Periodismo de Datos?

Creo que es el trabajo de poder encontrar historias, que muchas veces están presentadas directamente en los datos y es solamente traducirlos, es mostrar esa cantidad de números que están en una base de datos y presentarlos en un gráfico directamente; eso es Periodismo de Datos. Pero muchas veces es poder encontrar historias ahí, qué hay detrás de ese presupuesto que aumentó en un 50% de un año para otro en San Luis, eso que por ahí solamente es una barrita en el gráfico seguramente detrás de ello hay una historia, hay un cambio de políticas, hay algo que afecta a la población en sí; creo que son esas dos cosas. Poder traducir los datos de una manera simple y poder leerlos para ver si hay algo detrás, creo que eso es lo más divertido por ahí desde el lado del periodista, poder encontrar en esos datos información que se tuvo que brindar sí o sí y que está en manos del periodista poder encontrarle la vuelta si es que esos datos significan algo más de lo que realmente se están mostrando.

Creo que el trabajo que hay para hacer es divertido, creo que cambia muchísimo el juego, genera un nuevo tipo de periodismo y un nuevo tipo de periodista. Deja de ser el periodista que está solo en su redacción, encerrado en su oficina, buscando, llamando por teléfono, averiguando historias o generando párrafos inmensos para publicar una columna. Cambia a este tipo de periodista nuevo que está sentado delante de un montón de bases de

datos, de planillas de Excel, corroborando esa información con noticias que pasaron o con información que le hayan dado y que tiene que trabajar en conjunto con un montón de otras personas, creo que cambió totalmente la visión de lo que termina siendo un periodista.

El dato en sí o el número en sí puede mostrar algo que realmente después no se refleja, por eso hay que *matchearlo* con la realidad y con cosas que estén fuera de los datos, pero para mí el principal cambio es ese, cómo se transformó la idea o el rol del periodista en sí.

22. ¿En qué año se fundó Porven?

Nosotros empezamos a fines del 2014. En diciembre empezamos las primeras pruebas, empezamos a cargar algunas noticias a modo de ejemplo, empezamos a cargar los datos, a purificar todo eso que sabíamos que si no lo organizábamos bien desde un principio, en un futuro no íbamos a poder utilizar esos datos. Fue muy importante para nosotros superar esa primera etapa de prueba, empezar a cargar información. Lo trabajábamos en un grupo cerrado de personas en las que nos compartíamos las noticias dentro de Porven. Durante el 2015 fue que ya lo subimos online y lo pusimos accesible para todos, y ahí sentimos que ya habíamos armado una estructura con toda la información y la forma en que podíamos guardar y mostrar información, y que después en un futuro nos iba a servir para reutilizarlo.

23. ¿Cómo surgió la idea de hacer Porven?

Básicamente surgió medio como un capricho; particularmente yo y un grupo de personas con las que trabajo, siempre nos quejábamos de cómo se generaba el periodismo, más que nada en los diarios de Rosario: Rosario2, La Capital. En su momento, en el 2014, realmente los productos digitales que tenían eran muy malos, inclusive varios diarios a nivel nacional, las cosas que hacían a nivel digital eran muy malas, entonces todo el tiempo era una constante queja de “cómo puede ser que esto sea el medio que leen todos los rosarinos”. Y de un día para otro surgió la idea de ver cómo lo haríamos nosotros si tuviéramos que hacerlo, sin por ahí la idea ambiciosa de ser “el medio de rosario”, sino para ver cómo lo haríamos nosotros si tuviéramos que mostrar toda esta información que se genera en Rosario. Y así fue como empezó, y desde el día cero sabíamos que el Periodismo de Datos iba a ser una de las ramas principales de cómo íbamos a mostrar las cosas, porque iba muy de la mano con nuestra idea y filosofía general que es informar de una manera mucho más simple y que sea más fácil consumir la información.

24. ¿Tenían conocimientos previos de Periodismo de Datos o los fueron adquiriendo?

Teníamos conocimiento. Yo no vengo del palo periodístico, soy del palo tecnológico, entonces fue una incursión desde cero en el periodismo y fue desde el Periodismo de Datos. Desde que surgió la idea de empezar con el proyecto me empecé a empapar bastante en ese tema y obviamente cuando llegué al Periodismo de Datos descubrí que eso nos iba a ayudar muchísimo, así que fue un descubrimiento paralelo a la idea del proyecto.

25. ¿Tomaron como referente a La Nación Data? ¿Conocían el proyecto?

¿Piensan hacer algo parecido?

Sí, nosotros cuando empezamos el proyecto el diario La Nación no sé si ya estaba con ese equipo de La Nación Data, pero sí estaba haciendo muchos proyectos de Periodismo de Datos, y fue su forma de trabajar uno de los referentes que nosotros tomamos a nivel nacional. A nivel internacional había otros, The Guardian y algunos de Estados Unidos. Nuestra idea es a futuro, si podemos encaminar mejor el medio, tener un equipo directamente que trabaje sobre eso. Hay muchísima información para analizar de la Municipalidad de Rosario y del Gobierno de Santa Fe, los medios de Rosario no lo están haciendo o lo están haciendo de forma muy *light*, de forma tradicional, y creo que un equipo de Periodismo de Datos puede ser un eslabón entre esa cadena de datos que quiere liberar la Municipalidad y las personas que buscan consumirlos. Así que nos encantaría poder especializarnos más en eso.

26. ¿Planean seguir haciendo proyectos de Periodismo de Datos?

Sí, es más, tenemos algunos ahí en el *localhouse*, es decir, guardados de forma local; tenemos algunos de accidentes de tránsito que estuvimos analizando, el tema de mapas de cortes programados nos da a nosotros nuevos datos también. Nosotros hoy podemos generar un mapa de calor con las zonas de Rosario donde más se cortó la luz, entonces esa nueva información que se genera puede dar lugar a un gráfico, a una nota. Tenemos que seguir con esto porque sentimos que es realmente un ámbito donde nosotros nos podemos diferenciar del resto de los medios de la ciudad.

Entrevista a Ezequiel Clerici – Hacks/Hackers Rosario

21 de Junio de 2017

Entrevistador: Eugenia Michiels

Lugar: Rosario, Santa Fe

1. ¿De qué se trata Hacks/Hackers y cuándo comenzó el capítulo de Rosario?

Hacks/Hackers arranca en Argentina en el año 2011 a través del capítulo de Buenos Aires, con un grupo de Mariano Blejman, que era editor de la sección de cultura digital de Página/12, Martín Sarsale que es un emprendedor y un desarrollador de software que hoy tiene a su cargo el proyecto que se llama Properati, y Guillermo Movia que es un comunicador social y que trabajó durante muchos años como community manager de la Mozilla Foundation con todo lo que es el proyecto de Firefox. Eso es Argentina. Previamente en Estados Unidos nace a partir de la iniciativa de tres personas, arrancan este proyecto como una forma de experimentación por fuera de lo que eran sus entornos laborales profesionales, en principio porque uno trabajaba en el New York Times y el problema que tenía es, como en toda estructura tradicional de periodismo, es muy complejo lo que es la cuestión de innovación. Básicamente lo que encontraban es que no tenían un espacio de desarrollo dentro de sus lugares de trabajo, y deciden armar esto, sobretodo basado en una cultura que tiene mucho que ver con la cultura de Internet, esta cuestión de laburar interdisciplinariamente, que venga gente de distintos ámbitos a trabajar cuestiones relacionadas con medios de comunicación. La idea era pensar el futuro de los medios de comunicación, cómo íbamos a contar historias a partir de este momento. ¿Qué es este momento? Es un momento donde tenemos a disposición herramientas tecnológicas, tenemos a disposición una enorme cantidad de datos y de información dura que antes no teníamos, no es que no existía sino que simplemente no estaba a disposición como está ahora, y tampoco vos tenías la posibilidad de organizarla de manera sistematizada a través de software de computadora, o sea, periodismo computarizado. Ese es un poco el gen, después obviamente los capítulos crecen, tienen autonomía, cada uno funciona con su propia lógica, algunos están más orientados al periodismo, otros están más orientados al emprendedurismo, a proyectos que tengan que ver con un fin económico. En sí mismo no es una agrupación o una fundación a nivel internacional que diga “no tenemos fines de lucro”, es más, cada uno se cobija dentro del nombre de Hacks/Hackers con una lógica que es el periodismo en unos casos y en otros no, y a partir de ahí se trabaja.

El capítulo nuestro arranca en el año 2013 y nace con esta impronta. Yo, como vengo del periodismo y soy uno de los fundadores del capítulo, mi impronta es plenamente periodística, en un ambiente como Rosario donde innovar o tratar de trabajar este tipo de

temáticas es muy complejo, complejo por dos cosas: La primera es que no hay presupuesto, ese es el modelo de negocio de acá, no tener presupuesto y no se puede invertir. La segunda, con una lógica si vos querés reticente desde los sectores que más lo trabajan en el medio, editores, jefes y demás, muchos con una idea de tratar de innovar pero tampoco de una cosa demasiado fuera de lo normal.

2. ¿Con las herramientas que se encuentran en internet se puede hacer Periodismo de Datos?

Sí, pero es limitado. Yo creo que podés hacer más periodismo digital que Periodismo de Datos. Básicamente por que vos lo que necesitás también ahí es la mirada de un equipo interdisciplinario de trabajo, necesitás un programador, un programador que piense como programador con una mirada puesta en medios de comunicación. Vos podés visualizar un montón de datos con herramientas gratuitas o con herramientas libres, ahora lo que no podés hacer es tener un nivel de calidad como lo que tiene ProPublica por ejemplo, lo que tiene The New York Times, generar una *landing page* en donde toda una historia se cuente y se desarrolle desde una mirada integral, con esa estética. Podés tener igual un proyecto que use herramientas libres, que de hecho las hay, con una mirada integral de ese estilo, pero siempre vas a necesitar en un punto un programador que arme la estructura de lo que vos necesitás, a veces con motores libres. Esa también es otra cuestión, que es muy arraigada en lo que viene con la cultura de Hacks/Hackers, usamos software libre, generamos herramientas de software libre para que los demás las puedan tomar, utilizarlas, mejorarlas, contamos cómo son los procesos de trabajo, abrimos todo el proceso de producción de investigación y demás, contamos cuáles son las dificultades, contamos cuáles son las potencialidades, cuáles son los fuentes; porque la idea es justamente que esto sea lo más abierto posible. En eso se basa mucho en la idea de software libre, pero también del conocimiento abierto, que mucho tiene que ver con la lógica que tiene Firefox que tiene Duck Duck Go o la Electronic Foundation.

3. ¿Por quién está conformado el equipo de Hacks/Hackers?

Mayoritariamente somos periodistas o comunicadores sociales con un programador que va y viene, y lo que le falta, en particular, o lo que nosotros siempre creemos que es un déficit de Hacks/Hackers Rosario, es que no logramos atraer diseñadores gráficos. El núcleo duro somos unas cuatro o cinco personas: Leticia Leticia Giaccaglia que trabaja en

la Universidad Nacional de Rosario y también labura en la Municipalidad de Rosario en la parte de comunicación, Ivana Mondelo que también es comunicadora y trabaja en la Municipalidad, Sebastián Sorribas que es comunicador, trabaja en política y a su vez es el representante de Duck Duck Go en Rosario, Ezequiel Clerici que labura en Rosario3.com y en 5RTV, Rubén Acevedo que es programador independiente freelance, pero él va y viene, no tiene un rol tan decisivo, y Lucía Cuffia que es periodista y trabaja en PuntoBiz y como freelance. Ese es más o menos el grupo del núcleo duro, como verás tenemos un déficit de programadores en lo que es la toma de decisiones.

4. ¿El capítulo de Rosario depende del de Buenos Aires?

No. Es independiente, es como por una lógica propia de peso que tiene Buenos Aires, no solamente por la ciudad en sí misma sino porque es el capítulo con más miembros activos a nivel global. Eso tiene una explicación, y es que la Media Party influye muchísimo, porque para que vos puedas participar de la Media Party tenés que ser miembro de Hacks/Hackers Buenos Aires.

5. ¿Si sos de Rosario podés ir a la Media Party?

Sí. No hay organicidad, no hay una estructura piramidal, no hay una estructura jerárquica, las jerarquías se establecen de forma natural y se da en cada capítulo por los miembros que están más activos a lo que están menos activos.

6. ¿Cómo trabajan los periodistas, programadores y diseñadores en conjunto dentro de Hacks/Hackers?

Si querés saber cómo funciona la estructura de una mini redacción de Periodismo de Datos se puede pensar desde un proyecto, un proyecto de iniciación y finalización; la diferencia con un hackatón es que un hackatón es una jornada de trabajo intensa, se parte de un punto cero y se llega a un punto a,b,c, a una serie de puntos que después se tienen que continuar y seguir por fuera de lo que es un hackatón, por lo tanto ya queda dependiendo de ese equipo de trabajo. Inclusive Hacks/Hackers no tiene ninguna responsabilidad sobre ese equipo, lo único que nosotros hacemos es generar el espacio, el lugar y la propuesta para que venga gente a trabajar, después los equipos que se formen o la lógica que se arme ahí adentro es propia de cada equipo. Inclusive proponemos proyectos a sabiendas de que a veces es muy difícil que la gente venga con proyectos propios, con ideas propias, es una lógica muy de acá, de Rosario, que no se da en otros capítulos. Volviendo a lo que es el

trabajo en sí mismo, básicamente lo que se establecen son roles que son muy similares a los de un proyecto tecnológico habitual. Hay un líder de proyecto, un Project Manager, que en este caso sería un editor, la mirada periodística, si es un trabajo periodístico, que establece determinadas pautas de comunicación, tecnólogos que en este caso van a ser los desarrolladores y diseñadores, y después el equipo de periodistas que trabaje simplemente lo que es la parte de recolección de información que no se puede hacer de forma automatizada o que necesita interpretación. Después, lo que es el planteo de la plataforma en dónde se va a mostrar, es un laburo integral entre editor, programadores y diseñadores, en donde todos tienen una igualdad a la hora de opinar, obviamente la premisa de lo que es contar la historia la va a tener el editor. Siempre hablando de un caso periodístico.

No hay una estructura o una receta, la única receta que tenemos es que somos equipos multidisciplinarios de trabajo en donde el rol del periodista, la centralidad del periodista como la conocemos los que ejercemos la profesión, no existe. Es otra lógica.

7. ¿Hacks/Hackers Rosario tiene algún proyecto de Periodismo de Datos finalizado?

En HacksHackers Rosario tenemos varios, el que mayor repercusión tuvo y que llegó a estar en un medio de comunicación es el Mapa de Homicidios Dolosos de Rosario del año 2013. Se publicó en La Capital.

Básicamente lo que se hizo fue que se consiguió, a través de una fuente del Ministerio de Seguridad, una planilla en donde estaban todos los homicidios que se habían cometido en el departamento Rosario con nombre y apellido de la víctima, tipo de arma, edad, calle en donde murió, juzgado en donde cayó el caso, comisaría a la que perteneció. En base a esa información lo que nosotros hicimos fue limpiar todo ese dataset, que era desordenado, en el sentido de que la policía o los ministerios tienen una forma de hacer ese tipo de Excel no pensando con una lógica de apertura de datos, sino como una forma de registro. Entonces nosotros lo que le dimos fue esa estructura que necesitaba para ser leído sin ningún tipo de problema por un software de computadora, el cual se creó a medida con una licencia GPL, que es justamente de software libre, donde se mapeó toda esa información, se georeferenció cada una de esas muertes en donde estaba en base a las direcciones que se daban. Ese laburo se hizo de forma manual en una etapa, y de forma automatizada en otra, eso es porque era el primer trabajo a gran escala que hacíamos, y a medida que vas haciendo un primer proyecto te vas dando cuenta de todos los errores que vas teniendo y los vas

solucionando, los vas automatizando, los vas mejorando en base al error. Eso también es una lógica muy de laburar esto, aprendés haciendo. Como es algo nuevo se aprende haciendo, no hay otra manera, mucha gente que viene ya del palo de programación está muy acostumbrada a la prueba y al error, a trabajar de esa manera, y mucha de esa lógica se trasladó al momento de lo que es la digestión de los datos, el procesamiento de datos, como vos lo masticás a ese dato, que en eso los programadores tienen mucha más experiencia.

En base a ese dataset que habíamos conseguido, que tenía toda esta información, por una cuestión lógica de trabajo de que éramos un equipo de dos personas, un periodista y un programador (Pablo Cuadrado que hizo toda la parte de programación y de diseño); lo que hicimos fue, a medida que íbamos queriendo llegar a hacer algo, por ejemplo, queríamos, georeferenciar todas las direcciones, lo hicimos a través de un sitio web que decía coordenadas de latitud y longitud, pero era una cosa de locos tener que hacerlo una por una, pero lo hicimos así. Después descubrimos de que hay un código que se conecta con el API de Google Maps y que vos le tirás el dataset y te devuelve las direcciones, longitud y latitud, y te lo hace en 35 segundos; nosotros estuvimos meses haciendo eso, pero bueno, hacer eso durante meses te obliga a pensar una manera de hacerlo más fácil, porque al fin y al cabo lo que queremos todos es hacerlo en el menor tiempo posible y de la forma más fácil posible.

Esa lógica nosotros la fuimos aprendiendo en el hacer, en trabajar con esos datos, porque, por ejemplo, en un momento dijimos “bueno, ¿cómo lo vamos a mostrar?, tenemos todos los crímenes diseminados por la ciudad, los tenemos georeferenciados, ¿cómo hacemos para empezar a analizar esto? Bueno, los tenemos que dividir. ¿Qué tenemos? Tenemos las comisarías”, entonces mapeamos las comisarías, hicimos el polígono de la comisaría y del lugar que abarca, y ahí teníamos centrados la cantidad de homicidios. Entonces a partir de ahí ya empezamos a saber que las comisarías que más crímenes tenían eran todas las que eran de la herradura que hace de la zona sur y el noroeste. En esa franja teníamos la mayor cantidad de crímenes. Ya los teníamos referenciados por comisaría, entonces pensamos “¿qué zona es?” Porque esa después fue la otra, ¿qué distrito es? Para saber qué distrito era tuvimos que poligonizar esos distritos, nos tuvimos que tomar esa tarea que no estaba pensada al principio. Después decidimos que cada uno de esos puntos tenía que llevar el nombre, el apellido de la persona, la intersección de la calle donde sucedió y el tipo de arma que tenía, y es más, buscamos de los 214 crímenes que se habían

cometido en Rosario el registro en los diarios de la ciudad, encontramos cerca del 80 y 90%, y la mayoría los encontramos en La Capital. Eso nos hizo llevar a pensar en buscar a la gente de La Capital para proponerles de mostrar esto en su sitio web. Nosotros así pudimos tener una parte en donde contábamos todo el proceso.

Ya en el año 2013 sabíamos que en Rosario había un problema gigantesco en armas, que la población podía acceder a armas de fuego calibre 9mm, que es el calibre que usa la fuerza de seguridad o el ejército, que un civil no puede utilizar ni siquiera yéndola a comprar a una armería; tenés que tener una serie de requisitos y permisos para poder hacerlo. En principio eso, que eran armas reglamentarias que usaba la fuerza de seguridad y el ejército. Por lo tanto ahí ya había una hipótesis de conflicto, que alguien estaba vendiendo armas de guerra a las poblaciones que vivían en asentamientos irregulares. Porque también se daba que eran en barrios de marginalidad social donde sucedía la mayoría de los crímenes.

Después tenés la otra historia, que es la munición. Para vos comprar municiones también tenés una serie de requisitos, y se accede muy fácilmente a municiones. Años después saltó lo de Fray Luis Beltrán, de que faltaban dos millones de balas.

Después tenías la historia de la cuestión social. Por un lado la gente que tenía un problema interpersonal antes lo solucionaba cagándose a trompadas o de otra manera, ahora directamente va, busca un arma y le mete un tiro. La mayoría de la gente que se mataba, en general, no siempre, pero había un alto nivel o un alto índice de familiaridad; era gente que se conocía entre sí. La mayoría de esos crímenes no estaban atados a cuestiones de narcotráfico, que eso también era un dato muy importante, porque vos tenés que recordar que en 2013 había toda una estigmatización con el narcotráfico, se rompió un mito. Mirá todo lo que pudimos hacer en base a ese laburo, pero digo, ¿cómo lo hicimos? Sistematizamos un montón de información, que a simple vista no te decía nada, la volcamos en un mapa, la dividimos por zonas, por regiones, y empezamos a analizar en base a eso, que es lo que hace Análisis Criminal.

Nosotros hicimos una alianza estratégica con La Capital para que eso se publicase en el portal de diario y estuvo ahí como 3 meses.

8. ¿Y eso se visualizó con el mapa y con alguna nota basada en esos datos?

Sí, Hernán Lascano hizo una nota que acompañó al mapa, y nosotros hicimos la otra nota donde dimos nuestro análisis de los datos, que nosotros hacía casi medio año que los

veníamos estudiando. Y lo otro que te permite este tipo de visualización de grandes cantidades de datos es tomar dimensiones y empezar a encontrar historias, que de otra manera te costaría más verlas, no voy a decir que no las ves, simplemente que te costaría más; y lo otro, que es lo fundamental, y que es una condición *sine qua non* del periodismo, es que vos tenés que hablar con pruebas, tenés que hablar con hechos, tenés que hablar con datos que sean concisos, para darle credibilidad a tu artículo. Creo que es un poco lo que hace el Periodismo de Datos, que en realidad el Periodismo de Datos no es nada nuevo, es algo que ya se viene haciendo de la época de Napoleón Bonaparte. No es algo nuevo, pero sí viene o surge en un momento que se dan dos cuestiones bien claras, una crisis muy grande de credibilidad del periodismo, que es a nivel global, no se da exclusivamente en Argentina sino que se da en todo el mundo, se duda muchísimo de los medios de comunicación y de lo que informan, por una multiplicidad de razones, que tiene que ver con la masividad de medios que hay, con las redes sociales, con que cualquiera dice cualquier cosa, cuando cualquiera se transforma en un interlocutor válido, es como que se fue haciendo más líquida la credibilidad, porque hay un montón de medios de comunicación que informan, de ahí las fake news y etc.

Tenés dos grandes ramas, una que es la falta de credibilidad, entonces en base a eso vos tenés que empezar a ver de qué manera mejorás eso, y tenés a tu disposición una enorme cantidad de información que si la podés sistematizar y la podés aislar, podés generar nuevas historias, lo que pasa es que eso implica tiempo, tiempo que los medios de comunicación hoy no tienen, y no tienen porque eligen no tenerlo.

Por el otro lado tenés que internet, y todo lo que es la lógica de internet, está absolutamente masificada. Hoy está inmersa en todo lo que se hace hoy productivamente, los medios de comunicación no escapan a esa lógica, y muchos de los periodistas que hoy toman la bandera del Periodismo de Datos son gente que son, o nativos digitales, o millennials, que se criaron en una concepción del mundo que es completamente distinta a la que tenían los viejos periodistas, de un sistema analógico. Entonces, mucha de la lógica de internet, de compartir, de trabajar de forma multidisciplinaria, de abrir los procesos, una lógica que está mucho más inmiscuida con la lógica hacker, con la lógica misma de Internet.

También el Periodismo de Datos entra con fuerza después de lo que fue la explosión de Wikileaks, que Wikileaks es un paradigma de lo que es trabajar con datos; ellos no

puntualmente hacían visualizaciones, pero investigaban a través de periodismo computarizado, por ejemplo todo lo que fue los cables de las embajadas de los Estados Unidos, se trabajaba con software, eran tantos cables y en formatos muy específicos que utilizan los Estados, particularmente para compartir información, que había que sistematizarlo de alguna manera. Entonces, como que se juntan esas dos lógicas: una crisis de credibilidad del periodismo por un lado, y una cultura de internet con periodistas que tienen entre 20 y 40 años, entre nativos digitales y millennials, por otro; el diferencial no está en tener los datos y ocultarlos, el diferencial es qué vos haces con esa información, y tal vez si vos la abrís vas a conseguir más gente que se sume a ese proyecto y que permita una mayor apertura.

9. ¿Con qué bases de datos trabaja Hacks/Hackers Rosario? ¿Tiene una base de datos propia?

Ese es un proyecto que estamos tratando de generar, esto de tener datasets ya curados y propios a disposición para que la gente pueda tomarlos y hacer proyectos, lo que pasa que eso implica tiempo y trabajo, y eso si no está remunerado tiene determinadas limitaciones, que es la limitación de laburar en tu tiempo libre.

10. ¿Qué herramientas utilizan? ¿Utilizan el scraping?

Sí, usamos la base de Scaper y después es lo que vamos encontrando en el momento, algunos programadores diseñan sus propios scrapers a medida para buscar X cantidad de cosas, por ejemplo, para buscar tweets tenés una forma de generar scraping en HTML, después más que nada es programación de cero, se crea una herramienta de cero en base a código que podes forquear, se utiliza mucho D3JS, que es una librería de HavaScript. Después hay como una serie de repositorios externos, muy grandes, se labura mucho con mapas abiertos Open Street Maps, CartoDB, y después un montón de otras herramientas que son del ámbito de lo libre y que ya están pre-hechas, que lo único que tenés que hacer es cargarle la información.

11. ¿Trabajan con algún medio de comunicación? ¿Dónde publican las investigaciones?

Cuando tenemos algo listo lo que hacemos es, yo, por ejemplo, publico notas en Rosario3.com, ahí hago un laburo yo más de relaciones públicas, porque conozco gente en distintos medios, entonces les cuento lo que hay dando vueltas, qué hicimos, si lo quieren

publicar, si lo quieren tener en el sitio web o demás y se generan alianzas. Pero se hace todo a medida, en base a lo que se haga; si es una aplicación hacés una nota, contás que esa aplicación existe o tratás de que salga en algún medio de comunicación como fue la última que hicieron en el último hackatón que fue una aplicación para identificar vehículos habilitados de transporte público a través de la patente, entonces salió una nota en 5RTV.

En su momento, cuando fue el debate presidencial previo a la elección de 2015, hicimos una herramienta que lo que hacía era mapear todos los tweets que se iban generando durante el debate, las dos horas de debate. A eso le dimos manija por Rosario3.com, como cuando otros chicos por afuera de Hacks/Hackers Rosario hicieron Geografía Política, que lo que hacían era un mapeo de cómo había votado la sociedad rosarina en las elecciones de 2015. Eso se publicó, se hizo un meetup en donde se los invitó a que cuenten cual fue el proceso y demás. Se va tratando de buscar, lo que sí no hay es un interés decidido de los medios de comunicación locales de darle lugar a este tipo de cosas.

12. ¿Desde Hacks/Hackers Rosario tienen la idea de crear algo como el blog de La Nación Data?

Eso implica trabajo, implica tomarte el tiempo de escribir un blog, de actualizarlo, y la verdad que nadie tiene el tiempo como para sentarse a hacer eso. En un momento se hizo. Después vos también lo que necesitás es tener un dinero disponible para tener un dominio, un servidor, un mantenimiento. Ya implica toda una logística que cuando vos no trabajás por plata es más complejo. Tampoco tenés ningún medio de comunicación acá dispuesto a apostar a esto, más allá de Porven, que lo hace a pulmón y lo hace Horacio Bella.

Para mí una de las funciones primordiales del Periodismo de Datos es brindar un servicio público. Le tenemos que dar a la gente algo que le sirva; porque también lo que sucede mucho en el Periodismo de Datos es que se generan herramientas hiper complejas, cargadas de información, que vos las ves y son hermosas, pero al lector común no le dice nada, hasta lo agobia. Y por eso también se está volviendo más a un periodismo de investigación, apoyado en un Periodismo de Datos que no necesariamente se basa en aplicativos web que sean demasiado complejos.

13. ¿Qué productos de Periodismo de Datos realiza Hacks/Hackers Rosario? ¿Cada cuánto tiempo se realizan hackatones?

Mapas principalmente, visualizaciones y análisis en base a eso. El fuerte nuestro siempre ha sido mapas.

Hackatones hacemos una o dos al año como máximo, también eso depende de la gente, si la gente se copa los hacemos, pero organizar un evento lleva mucho laburo. Sí lo que hacemos mucho es invitar todo el tiempo a la comunidad a que plantee temas, a que plantee meetup,s a que plantee hackatones, y que obviamente se lo pongan al hombro con toda nuestra logística. Nosotros proveemos los lugares y los medios para que eso ocurra. Nosotros lo que queremos hacer es un espacio de *networking* en donde vos vengas a mostrar un proyecto, a trabajar, a hacer lo que se te cante, y encuentres en esa red que nosotros armamos la gente indiada con la cual puedas llevar adelante tu proyecto, después tu proyecto es tuyo, no es de Hacks/Hackers, a no ser de que salga del núcleo de Hacks/Hackers, de los coorganizadores. Ahora estamos haciendo un laburo que se llama Quién es Quién, que lo que intenta a hacer, en un principio, es una gran base de datos con los directivos de las principales empresas de Argentina, para después cruzar esa información con organigramas del Estado para ver si hay conflicto de intereses, pero sobre todas las cosas crear un sitio web abierto para toda la comunidad en donde vos puedas ir a buscar una empresa y rápidamente te devuelva quiénes son el CEO, los directores ejecutivos, quién se encarga de comunicación, cuánto factura, cuántos empleados tiene, cuál es el número de CUIT, para que vos tengas esa información a disposición. Ese proyecto nace del núcleo de Hacks/Hackers Rosario, ese proyecto va a ir con la firma de Hacks/Hackers Rosario, pero en el medio se han hecho un montón de aplicaciones que sonde la gente que lo hizo en el marco de Hacks/Hackers Rosario.

Nosotros lo queremos es ser un espacio en donde cualquier persona que trabaje con periodismo, con el mundo digital, con emprendedurismo, con la programación, con el diseño web y le interesen los temas que nosotros planteamos o que plantea la comunidad, que sepa que en ese lugar, cuando vaya a un encuentro, se va a encontrar con toda gente que está en la misma, y tal vez eso genere oportunidades de trabajo, oportunidades de proyectos, etc.

14. ¿Cuál sería tu definición de Periodismo de Datos?

Es periodismo ayudado por computadora que aprovecha las grandes cantidades de datos que tenemos disponibles a partir de lo que es la naturalización de Internet y de que todo sucede a través de Internet, a que los gobiernos de a poco van aplicando filosofías de

apertura de datos, que en Argentina son extremadamente pobres y segmentadas. Y que lo que intentamos hacer es aprovechar todo eso basados en la cultura de Internet, para generar historias que sigan teniendo como meta o como objetivo el periodismo puro y duro, el periodismo tradicional, investigar en base a hechos reales, comprobables y verificables, y escribir un artículo que tenga todo eso con todas las voces a favor y en contra, y que al lector le permita tomar una decisión que mejore su vida, o que por lo menos le permita elegir con mayor claridad a la hora de tener que hacerlo.