



1. Small Data y Big Data: la evolución del periodismo de datos

“El universo está compuesto de energía, materia e información, pero es esta última la que lo hace interesante. Sin ella el universo sería una sopa amorfa. Carecería de las formas, las estructuras, los órdenes (...) que le dotan tanto de su belleza como de su complejidad”. Así comienza César Hidalgo su libro *El triunfo de la información* (2017) cuya propuesta es un viaje para descubrir las estructuras y órdenes de una nueva economía: los datos. Esa economía que ha impactado diversos quehaceres -o pequeños universos- entre los cuales está el periodismo de datos.

¿En qué medida los argumentos de un físico del MIT, y sus ideas sobre cómo formas y estructuras combaten la entropía, se relacionan con nuestra disciplina? En todo, me atrevo a contestar; siendo coherente con lo que digo en el libro *Números que hacen Noticia* (2015) y majaderamente repito en clases a mis alumnos: “informar es poner en forma”.

Para los data-reporteros las formas tienen mucho valor, en cada etapa de trabajo. Veámoslo con un ejemplo muy sencillo:

Desde el punto de vista de los datos las personas somos fuentes móviles no estructuradas. A donde sea que vayamos nos dirigimos con una carga de datos en la piel: edad, sexo,

nacionalidad, miedos, recuerdos, olvidos, expectativas, grupo de sangre, enfermedades, número de calzado, peso, colesterol, profesión, hobbies, historial educativo, registro de notas, ingresos económicos, deudas, pareja, exparejas, quizá amante(s) y un interminable etcétera.

No somos conscientes de todos los datos que portamos (como la presión arterial); a veces tenemos pruebas materiales (fotos de infancia, certificados de estudio) y otras veces los registros están en manos de otros (nuestras búsquedas en Google, registros de bienes raíces, deudas bancarias, etc.)

Visto así, el potencial de datos que ofrece cada individuo es muy rico, pero esconde un desafío que de no abordarse correctamente podría convertirse en un problema: estructurar los datos. Si lo hacemos mal podríamos terminar con un cúmulo de información inútil.

Aterricemos esto a algo de valor periodístico. Supongamos que necesitamos revisar datos de personas que trabajan como lobistas para detectar posibles conflictos de intereses y corrupción. Tenemos acceso a los registros de lobby en Chile gracias a la ley 20.730 (datos estructurados en tablas Excel) e información familiar de cada uno de ellos, a través del registro civil (dato en PDF, digital, no estructurado, no editable). Hemos recopilado archivos de prensa con apariciones públicas de estos personajes y datos sobre su vida; son diarios en papel (dato no estructurado, no digitalizado, ni editable). Finalmente, tenemos en agenda varias entrevistas con personajes que resultan relevantes para nuestra investigación.

Estructurar coherentemente los datos nos asegurará encaminar correctamente el análisis. La información resultante será valiosa como insumo y como guía, porque nos permitirá hacer preguntas más precisas en las entrevistas agendadas. Doble ganancia.

Las estructuras de datos en el periodismo siempre están relacionadas con preguntas de reporte. Mejores son las preguntas, mejor la estructura. Mejor la estructura, mejor el análisis. ¿Cómo podemos saber si las preguntas son mejores o peores? La respuesta reside en un criterio netamente periodístico: buenas preguntas son aquellas que me dirigen hacia el foco de la investigación.

Siguiendo con nuestro ejemplo, nos habíamos propuesto detectar posibles conflictos de intereses y corrupción de parte de personas dedicadas al lobby. A continuación, presentaré tres tipos de estructura. La primera responde a las siguientes preguntas:

- ¿Quiénes son las personas que se dedican al lobby en Chile?
- ¿Qué edades tienen?
- ¿Están casados?
- ¿Qué profesión u oficio tienen?
- ¿Registra acciones de lobby?
- ¿Cuántas acciones de lobby registra?

Las preguntas nos llevan a esta estructura:

Nombre	Edad	Casado(a)	Profesión u oficio	Registra acciones de Lobby	Cuántas
Pedro Fantiss	25	Sí	Periodista	Sí	4
Claudia Arbel	37	Sí	Abogada	Sí	3
Alex Netter	32	Sí	Administrador Público	Sí	5

La tabla resultante es una forma de ordenar los datos, pero muy inútil para nuestros fines ¿Cuánto nos ayuda saber la edad, si está casado, su profesión o cuántas acciones de lobby registra, para descubrir si ha estado envuelto en corrupción o conflicto de intereses? Nada.

Probablemente todos los temas contenidos en esas preguntas y la estructura resultante sean de interés, pero falta afinar la puntería. Veamos cómo se puede mejorar.

Los conflictos de interés y corrupción encierran conflictos de relaciones. Muchas veces se trata de acciones para beneficiar a un tercero relacionado, o la utilización de un tercero para obtener un beneficio de vuelta. Entonces, más que saber si está casado, nos interesa saber con quién; más que saber su profesión, nos interesa saber dónde ha trabajado; más que saber cuántas acciones de lobby registra nos interesa saber con quiénes ha tenido dichas reuniones y descubrir cuáles no registra.

La segunda tabla queda así

Nombre	Cónyuge	Empleos anteriores	Con quién NO registra reuniones
Pedro Fantiss	Amaya Foler	Minera tres cables	Ministro Sagrado Bozari
Claudia Arbel	Max Antiao	Petrolera DNAR	Ministro Alessio Pérez-Roca
Alex Netter	Juller Sato	Áridos "Granito Feliz"	Ministra Clara Valdenna

Esta tabla me acerca un poco más a lo que quiero descubrir. Es razonable pensar que alguien no registra una reunión porque algo quiere esconder, ya que la ley le obliga a hacerlo (en el caso chileno la ley obliga a las autoridades. Siguiendo este caso inventado, a los ministros Bozari, Pérez-Roca y Valdenna). Tener ese dato es útil, como también lo son los nombres de sus cónyuges. La historia de escándalos de corrupción está llena de cheques emitidos a nombres de anónimos esposos o esposas, asegurando que la plata llegara a la misma casa.

Aun así, la tabla 2 es una estructura incompleta. La podríamos complementar organizando por separados, datos de los cónyuges y las contrapartes del lobby no declarado:

Nombre	Empleos
Amaya Foler	Socia de B&A Consultores
Max Antiao	NO
Juller Sato	Universidad del Monte Verde

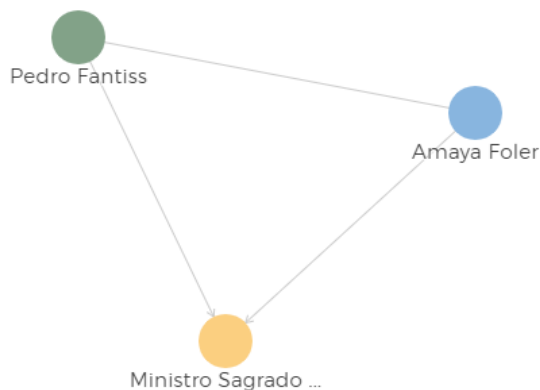
Ministro Sagrado Bozari	Socio de B&A Consultores
Ministro Alessio Pérez-Roca	Bellomundo TV.
Ministra Clara Valdenna	Aguas Claras S.A

La forma en que ordenamos los datos nos evidencia algo interesante.

El lobista Pedro Fantiss se reunió con el Ministro Sagrado Bozari sin que éste lo declarase. Bozari conoce de cerca a la esposa de Fantiss, Amaya Foler. Ambos son socios de B&A Consultores.

Para ofrecer nuestro hallazgo al público, tendremos que reestructurar los datos otra vez, porque las tres tablas por separado no aseguran lo que buscamos en el periodismo: que se entienda con claridad lo que queremos decir.

Probemos con un mapa de nodos. Al *clickear* cada nodo se accede a la información de los personajes:



El ejemplo anterior es bastante más simple que la vida real, pero sirve para instalar una idea fundamental en el periodismo de datos: se trata de estructurar y reestructurar una y otra vez; de ordenar, de dar con las formas correctas para que se presente ante nuestros ojos la revelación que convierte a la infoxicación en información: el dato valioso que vuelto a estructurar, pensando en el público, informa.

Dicen los mexicanos que la mejor forma de esconder un elefante en el zócalo es llenar el zócalo de elefantes. Tienen mucha razón. Podríamos agregar que la mejor forma de esconder datos valiosos es llenar de datos inútiles. Estructurar correctamente los datos a partir de preguntas bien formuladas y de valor periodístico, que apunten a un objetivo de investigación, es una manera de combatir ese problema. Una, porque el análisis de datos exige mucho más: estadística, programación, sentido del diseño, por ejemplo.

Y entonces, ¿Qué es periodismo de datos?

La moda y la escasa formación han puesto muchos pelos en la sopa del periodismo de datos a lo largo del tiempo, haciendo de su definición conceptual una bolsa donde cabe cualquier cosa.

Los que no quieren lidiar con programadores ni estadísticos lo definen únicamente como periodismo de investigación al alero de la ley de transparencia; los que no soportan una investigación de largo aliento dicen que solo se trata de meter datos interesantes en una nota de color; los ortodoxos de la estadística se quedan solo en la pulcritud del número, olvidando a los destinatarios no especializados.

Ofrezco mi definición: El periodismo de datos es periodismo de precisión que considera a los datos como fuentes entrevistables. No confía en éstos sin antes haberlos cuestionado y analizado con rigor metodológico. Es capaz de jerarquizar los datos valiosos y ofrecerlos a un público masivo aprovechando la riqueza narrativa que ofrecen los medios digitales. El resultado es periodismo de investigación para ecosistemas digitales.

En 2015 en la entonces Escuela de Periodismo de la Universidad Adolfo Ibáñez publicamos el libro "**Números que hacen noticia**".

Decíamos entonces

"¿De qué sirve que un gobierno libere montones de cifras de interés público, si lo hacer a través de pesados archivos en PDF difíciles de entender? ¿cuánta gente tendría el tiempo y la capacidad de interpretarlos y cruzar variables para descubrir, por ejemplo, si el dinero de los impuestos se gasta correctamente? El periodismo de datos se ha hecho cargo de acercar al público esos pesados archivos. Investiga qué quieren decir esos números, construye historias y las presenta a su audiencia a través de visualizaciones atractivas y de fácil comprensión. Periodistas, diseñadores y programadores trabajan en equipo y logran que ese cúmulo de cifras se transforme en algo inteligible"

En los 4 años que han corrido -y el agua respectiva bajo el puente- la evidencia muestra que, para hacer inteligible su trabajo, los medios de comunicación realizan tipos de periodismo de datos que podríamos agrupar tal como los géneros periodísticos tradicionales: informativo, interpretativo y de investigación

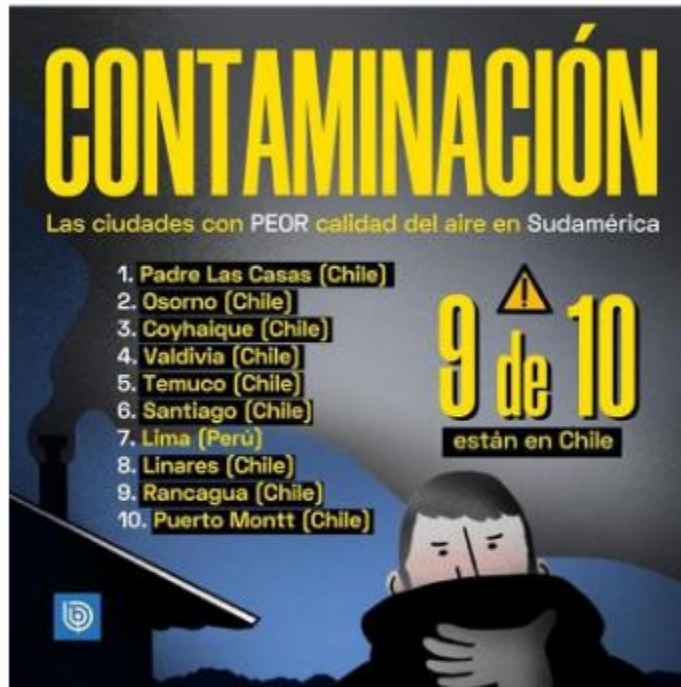
Periodismo de datos informativo

Toma datos de fuentes confiables, los estructura con sentido jerárquico y los ofrece al público a través de visualizaciones claras, precisas y concisas. En Chile **Radio Biobío** ha hecho trabajo muy interesante, aprovechando medios como Instagram, para ofrecer por esa vía sus contenidos al público. Datos que son números, hechos e hitos.

Tomándome del título de este capítulo, aquí hablamos de **small data**. Podemos realizar los análisis en forma manual, con herramientas como Excel, usando tablas dinámicas y filtros sencillos.



biobiochile · Siguiendo



4.974 Me gusta

biobiochile Cuesta creerlo, pero es verdad: 9 de las 10 ciudades más contaminadas en Sudamérica son chilenas.



biobiochile · Siguiendo



ALAN GARCÍA
Las claves de la investigación

- * Enero 2017:** Jorge Cuba, exviceministro de Comunicaciones de García es detenido por presunto cobro de coimas a Odebrecht para licitar la Línea 1 del metro de Lima
- * Marzo 2017:** García es denunciado por "delito contra la administración pública" por el mismo caso
- * Julio 2017:** Fiscal Sergio Jiménez confirma investigación contra García
- * Febrero 2018:** Jorge Barata, director ejecutivo de Odebrecht, declaró haber entregado US \$200.000 al partido aprista en 2006
- * Junio 2018:** Fiscal José Domingo Pérez abrió investigación preliminar por lavado de activos a García por caso Odebrecht
- * 17 de Nov. 2018:** Poder Judicial decretó 18 meses de arraigo nacional a García
- García pide asilo político a Uruguay
- 3 de dic. : Uruguay rechazó asilo
- * 17 de abril 2019:** **García se suicida** con un disparo en la cabeza antes de ser detenido



997 Me gusta

biobiochile Paso a paso, así se dieron los hechos previos al suicidio de Alan García, dos veces presidente de Perú y vinculado al caso Odebrecht.

- **Periodismo de datos interpretativo:**

Puede iniciar camino igual que el informativo, pero va un paso más allá. El conocimiento de estadística es muy necesario, pues implica cruces de datos, analizar muy bien las metodologías de quienes recolectaron dichos datos, saber qué es una ficha técnica, normalizar correctamente si es necesario.

El periodismo de datos interpretativo no solo se construye solamente con datos recolectados por otros, sino también con propios. Convierte declaraciones, situaciones y hechos en datos con estructura, para analizarlos. En esta categoría entra, por ejemplo, el *Fact-Checking*.

- **Periodismo de datos investigativo:**

Se ha desarrollado en Chile al alero de la ley de transparencia (20.285). Es el más robusto de los tres porque los integra a todos. Ahonda en temáticas complejas, de alta connotación y, por lo mismo, aspira a convertirse en un golpe periodístico.

Cuando hace gala de su naturaleza digital, es periodismo que integra muy bien RECOLECCIÓN + ANÁLISIS + VISUALIZACIÓN, una ecuación que en Chile es más bien escasa; pues, si bien hay periodismo de datos investigativo de excelencia, las narrativas son muy pobres.

En cuanto a lo investigativo y el uso eficiente de ley de transparencia con foco en lo relevante [CIPER Chile](#) es, sin duda, el mejor exponente. Se nota el peso y la experiencia de quienes le dan vida.

› TRANSPARENCIA

ACTUÓ COMO LOBISTA ANTE SIETE ENTIDADES PÚBLICAS MIENTRAS FUE ASESOR LEGISLATIVO

Senado contrató a abogado de Imaginacion aunque sus normas prohíben emplear a lobistas

24.05.2019

Por Nicolás Sepúlveda

TEMAS: Congreso, Imaginación, Lobby, Senado, Transparencia



Evolución

La masificación de la tecnología, la cultura open data y legislaciones de transparencia aprobadas en los últimos diez años permitido la liberación de datos antes guardaban celosamente los aparatos estatales, como presupuestos, sueldos de autoridades, gasto militar, entre otras. Para los periodistas esto representa el acceso a nuevas fuentes, a la vez que la promesa de una sociedad más abierta -ergo, democrática-; pero también el desafío de hacerse cargo de un dilema cada vez más pesado: ¿cómo lidiar con el abrumador cúmulo de datos disponible, para dar con lo relevante y prescindir de lo inútil?

En el manual publicado en 2015 nos hacíamos cargo de ese desafío desde una metodología que incluía estrategias de *wobbing* (uso de ley de transparencia), *crowdsourcing* (colaboración

abierta de personas), *scraping* (raspado automatizado de información) y el trabajo con plataformas de datos abiertos. Dichas temáticas, si bien son útiles para iniciar camino en esta disciplina se han vuelto insuficientes. Esa es la razón de este nuevo manual y el formato que hemos elegido para ofrecerlo al público: como una plataforma de contenidos digitales que permita actualizar atendiendo a los cambios acelerados en esta industria.

Un año después el caso *Panamá papers* puso sobre la mesa el trabajo con big data. La investigación tuvo como insumo, 11,5 millones de documentos. Mediante programación y estadística, periodistas conectados en todo el mundo identificaron patrones y consiguieron establecer perfiles para identificar el tipo de evasores que habían usado empresas de papel en paraísos fiscales, con el apoyo del estudio de abogados Mossac Fonseca en Panamá.

Tal cantidad de datos tuvo diversas formas de visualización para el público. Confieso haber celebrado una en particular: un juego de rol que permite al usuario descubrir cómo actuaron los evasores. ¿Por qué?, en 2015 pedí que la contratapa del libro fuera explícita al señalar: "el próximo golpe periodístico podrías verlo en un videojuego". Cuando presentamos el texto en octubre de ese año, ante prestigiosos periodistas chilenos hubo varios que me dijeron "¿te pegaste en la cabeza? Los juegos son una cosa y el periodismo es otra".

Insisto. Gráficos, mapas, juegos... son formas y de eso se trata: de poner en forma.

El periodismo de datos ha tenido un crecimiento importante en la última década, favorecido por transformaciones socioculturales, económicas, tecnológicas y políticas. Cambios que han favorecido la aprobación de leyes de transparencia, la irrupción de una ciudadanía empoderada, un contexto de cultura colaborativa, la masificación de dispositivos móviles como nuevos instrumentos -personalizados- de comunicación, la programación como un nuevo lenguaje universal y nuevas interfaces.

Marco el último punto de la lista: programación como algo fundamental para los data-reporteros de este tiempo. ya no basta con tablas dinámicas ni filtros en Excel (son un buen punto de partida, por cierto) pero es fundamental conocer lenguaje de programación: R y Python asoman como las opciones. Así lo han hecho notar, en el uso que dan en sus salas de redacción, medios como *The New York Times*, *Washington Post* y *The Economist*, pero mencionar algunos. Al principio un Notebook de código puede parecer el infierno, pero se ha ido generando un medioambiente interesante para buscar opciones de inteligencia colectiva: en GitHub, por ejemplo, los equipos de datos más consolidados ofrecen los códigos de programación de sus reportajes. El objetivo detrás de esto es apuntar a la reproductibilidad: es decir, que otros puedan hacerlo. Aprender por imitación podría ser una buena opción.

Cada elemento de este camino es una nueva forma: nuevas formas de fuente, de acceso a la fuente, de consumo, de interacción, de creación de contenido.

Si el periodista quiere ser el superhéroe que los relatos de ficción presentaron a través de Clark Kent y el joven reportero gráfico Peter Parker y que en el mundo real encarnaron los notables Woodward y Bernstein, ya no basta con ser un buen observador de la realidad ni las clásicas habilidades descriptivas. Ambas son importantes, sin duda, pero no bastan. Como no basta la sola observación para el médico y el astrónomo, el presente exige nuevas competencias a los profesionales de la información: análisis de datos, programación, ingeniería inversa y algoritmos debiesen ser parte de nuestro día a día.