

¿Qué es el periodismo de datos?

Esta disciplina periodística se nutre de otras muchas: de investigación, en profundidad, de precisión, asistido por computadora y analítico. En ella, se trabaja **con grandes volúmenes de datos**, se aprovecha al máximo la **visualización interactiva** y se incorpora al **programador** al equipo periodístico.

SANDRA CRUCIANELLI

Muchos colegas me preguntan qué es el periodismo de datos (PdD). No encuentran lógica en la denominación, ya que, desde siempre, los periodistas nos valemos de datos para producir contenidos periodísticos. Corresponde, en realidad, llamarlo “periodismo de base de datos” (PBD); pero, por uso y costumbre, predomina la denominación más corta.

- ¿Es periodismo de investigación (PI)?

Ciertamente, se usan técnicas propias del periodismo investigativo.

- ¿Es periodismo en profundidad (PP)?

También, dependiendo del tema que se esté analizando.

- ¿Es periodismo de precisión (PPr)?

En ocasiones, efectivamente, hay que recurrir a los métodos de investigación social definidos en la década de los 60 por

Philip Meyer para su aplicación en el campo del periodismo.

- ¿Es periodismo asistido por computadora (PAC)?

Casi siempre, porque tenemos que analizar datos y nos valemos de hojas de cálculo, así como de los procedimientos propios de esta disciplina. De hecho, creo que el PAC solo queda resumido a cuando manejamos volúmenes de datos pequeños y, para casos más grandes en cuanto a volumen de información, ha dejado de ser PAC para convertirse en periodismo de datos.

- ¿Es periodismo analítico (PA)?

También, ya que en general usamos métodos analíticos para llevar los datos a sistemas de información geográfica que nos permitan visualizar mejor el comportamiento de las variables que tenemos.

En realidad, lo que hoy se conoce como

Sandra Crucianelli es periodista especializada en periodismo de investigación y periodismo de precisión, fundadora de Sololocal.info, miembro del Consejo Asesor del Centro de Periodismo Digital de la Universidad de Guadalajara (México), docente del Programa de Educación Virtual del Centro Knight de la Universidad de Texas y del Centro de Medios de la Universidad de Florida (ambas en EE. UU.).

periodismo de datos –que, como ya dije, es verdaderamente periodismo de base de datos– involucra todo lo anteriormente citado, a lo que debemos agregar:

- 1.- Grandes volúmenes de datos, que muchas veces ni siquiera podían alojarse en una hoja de cálculo de Excel
- 2.- Visualización interactiva
- 3.- Incorporación del programador al equipo periodístico para desarrollar tareas que van desde la extracción de datos y la depuración de los sets de datos hasta el diseño de aplicaciones de noticias (*news apps*)

Entonces, si pudiéramos escribirlo a modo de fórmula, podemos decir que tenemos una suma de métodos conocidos a los que se añaden tres frutos de la innovación tecnológica:

PI + PP + PPr + PA + PAC + volumen de datos + visualización interactiva + programación = PdD o PBD
--

Para que comencemos desde cero y quede todo claro, las definiciones son las siguientes:

- **Periodismo de investigación:** se requieren tres condiciones necesarias para que un reportaje se precie de ser investigativo: que sea un tema de relevancia social, a lo que se añade que alguien pretenda mantenerlo en secreto y, finalmente, se requiere un trabajo adicional del periodista, ajeno a las prácticas cotidianas de cobertura informativa. Son los casos en los que se investigan asuntos relacionados con irregularidades en la administración del dinero público, como los casos de corrupción, cuando se

revelan situaciones de ineficiencia o abuso por parte de compañías privatizadas de servicios públicos y todo aquello que cause una afectación a la sociedad (por ejemplo, un caso de delito ecológico).

- **Periodismo de profundidad:** es lo mismo que el PI, solo que en este caso nadie intenta mantener en secreto lo que el periodista investiga. Así, por ejemplo, un estudio sobre las variables de pobreza con datos de los censos nacionales.

- **Periodismo de precisión:** formulado como disciplina –como ya apunté– por el profesor Philip Meyer, involucra el uso de métodos de la investigación social aplicados al periodismo tales como la investigación cualitativa (análisis de casos, grupos focales, análisis del discurso, etc.) y la cuantitativa. Esta última incluye la descriptiva (cuando se consideran todos los casos: por ejemplo, un censo entre legisladores) y lo que se denomina investigación inferencial, es decir, cuando se trabaja sobre un universo y, dada la dificultad para abordar a cada miembro de ese universo, se diseña una muestra representativa, a la cual se procede a poner bajo estudio. Es el caso de las encuestas de opinión o las electorales con fines predictivos.

- **Periodismo analítico:** analiza una realidad compleja, con el fin de lograr la comprensión del público. Combina algunos aspectos del periodismo de investigación y la información explicativa. Su misión es ofrecer pruebas basadas en interpre-

taciones de la realidad, explicándolas de un modo comprensible. Contextualiza un tema de fondo mediante la descripción de detalles históricos y datos estadísticos. El resultado es una explicación completa, destinada a dar forma a la percepción del público del fenómeno. Aspira a recoger datos dispares y hacer conexiones que no son inmediatamente evidentes, y es eficaz a menudo en el análisis entre los hechos y no en los hechos mismos. Utiliza como pilar de la visualización los llamados sistemas de información geográfica (SIG).

- **Periodismo asistido por computadora:** conocido como CAR (*Computer Assisted Reporting*) por sus siglas en inglés, se define como “cualquier proceso que utiliza la asistencia de computadoras durante la recolección y procesamiento de datos”. Ha basado su práctica en el uso de hojas de cálculo de Excel para el análisis de información contenida en tablas.

Los productos del periodismo de datos

Hay cuatro productos diferentes al menos, generalmente complementarios, que surgen de esta práctica:

- 1.- Artículos basados en datos
- 2.- Visualizaciones interactivas
- 3.- Conjuntos de datos abiertos
- 4.- Aplicaciones de noticias

- **Artículos basados en datos:** generalmente, son artículos cortos construidos a partir de grandes volúmenes de datos como los que podrían estar contenidos en bases de datos o conjuntos de datos diseñados

a partir de una o varias bases de datos (<http://www.lanacion.com.ar/1484852-los-millones-de-la-ape-como-se-repartio-en-2011-la-caja-que-era-de-moyano>).

- **Visualizaciones interactivas:** en general, son complementos de los artículos basados en datos, pero pueden ser contenidos en sí mismos sin necesidad de que exista un artículo referencial: en algunos casos, es suficiente un título, un subtítulo y una explicación a la audiencia sobre cómo leer los datos desde la visualización (<http://www.lanacion.com.ar/1546303-los-bienes-de-los-funcionarios-en-la-primera-news-application-de-la-nacion>).

- **Conjunto de datos abiertos** (conocidos como set de datos): los medios de comunicación no necesariamente se abocan a realizar trabajos periodísticos basados en datos. También pueden, como el caso de *La Nación* de Argentina, dedicar una gran parte de su esfuerzo a la apertura de datos públicos. En países en los que no existen leyes de acceso a la información pública, estas no funcionan como deberían, en los sitios webs del Gobierno hay algunos datos pero otros no, están ocultos bajo distintas capas de información o los formatos para mostrar los datos en la web no son accesibles (ver un dato en la pantalla no es lo mismo que acceder al mismo). Para que un dato sea accesible se debe mostrar en un formato reutilizable. Por ejemplo, un documento oficial subido a Flickr. En estas circunstancias, medios y organizaciones pueden dedi-

carse a la búsqueda, extracción y apertura de datos públicos, tal como *La Nación* hace a través de su catálogo de datos. Esto es, ya que hay datos a los que los Gobiernos o las organizaciones no permiten acceder –sea por la razón que fuere–, los medios de comunicación pueden incursionar en ese terreno, y así seguramente harán grandes aportes. Los datos pueden gestionarse o administrarse desde plataformas sencillas y gratuitas como Google Drive (<http://data.lanacion.com.ar/dashboards/5068/inflacion-y-precios/>).

- **Aplicaciones de noticias:** conocidas en inglés como *news apps*, son el fruto de la labor conjunta de periodistas y programadores informáticos. Muchas veces, el volumen de datos es tan grande que resulta imposible encontrar una noticia si no se diseña una aplicación que nos permita agrupar y analizar variables, por ejemplo, por localización geográfica, por fecha, por nombre de compañía, etcétera (<http://www.gastopublicobahiense.org/> - http://www.nacion.com/CustomerFiles/nacioncom/Generales/Subsitios/Investigacion/2012/propiedadessubvaloradas/props_diputados/index.html - <http://www.texastribune.org/library/data/government-employee-salaries/> - <http://projects.propublica.org/docdollars/>).

Sin cultura de datos abiertos

No existe una cultura de datos abiertos en la mayoría de los países de habla hispana. Sin embargo, lo que no hagan los

Gobiernos en materia de datos abiertos lo hará la comunidad, las organizaciones de la sociedad civil o los medios de comunicación.

El pronóstico no es nuevo. Un ejemplo vio la luz en 2010, cuando el diario *Texas Tribune* solicitó al Gobierno estatal el detalle de los salarios de los 674.000 empleados públicos, con la intención de diseñar una aplicación que permitiera a los ciudadanos consultar los haberes que se pagan con fondos públicos, incluidos maestros de escuelas y profesores de universidades públicas (<http://www.texastribune.org/library/data/government-employee-salaries/>).

Para hacer periodismo de datos se puede prescindir de muchas cosas, ya que hay herramientas digitales gratuitas que todos los periodistas podemos usar para sortear casi cualquier obstáculo, pero de lo que no podemos prescindir es de los datos. Son una condición necesaria sin la cual no podemos siquiera comenzar a diseñar hipótesis periodística alguna.

En muchos casos, el formato en que se visualizan los datos en las webs no es reutilizable. Verdaderamente, los periodistas podemos resolver estos escollos aprendiendo técnicas específicas. Pero ¿qué ocurre con los ciudadanos? ¿Acaso los contribuyentes no tienen derecho a un mejor acceso a los datos directamente relacionados con el dinero público?

La información con atraso, desordenada, aún desactualizada o en formatos inadecuados se puede leer en pantalla

ciertamente, pero visualizar datos no es lo mismo que acceder a ellos. Es el clásico ejemplo de una imagen convertida en un PDF en el que se lee el detalle de una tabla, como la ejecución de un presupuesto. Habría que recurrir a una técnica de reconocimiento óptico de caracteres (OCR, por sus siglas en inglés) para extraer los datos; aunque estos programas no siempre resultan eficientes y no queda otro recurso que la carga manual de datos, lo cual complica enormemente la tarea de todo reportero, en especial si el contenido de la tabla que queremos procesar es extenso.

En el filme *Ojos bien cerrados*, dirigido en 1999 por Stanley Kubrick poco antes de su muerte, el eje argumental versa sobre dos sentimientos concretos: la incertidumbre y la insatisfacción. Por poner un ejemplo, en Argentina, sin los componentes del *thriller* que rodean aquella oscura historia, 103 años después, ciudadanos y periodistas nos encontramos con *Datos bien cerrados*. Ante ellos, experimentamos incertidumbre y nadie puede sentirse satisfecho con un escenario en el que hay que tener habilidades específicas, cuando no pagar por recursos adicionales, para acceder a datos legítimamente públicos.

Cómo armar un set de datos

Los datos ya han sido recuperados de la web o descargados en el PC. Tal vez, el periodista haya tenido que recurrir a la entrada de datos manual, ya que esto es frecuente cuando los documen-

tos, mediante petición, son entregados en formato papel y no resulta posible la extracción de la información y su posterior reestructuración en tablas, considerando que no siempre funcionan con eficiencia los programas de reconocimiento óptico de caracteres.

En periodismo de datos, consideramos que un set de datos es una colección de elementos de datos agrupados, que permiten su fácil registro. En ellos, los datos se encuentran organizados en variables, por lo general, medidas a lo largo del tiempo y en soportes descargables.

Visualizar datos no es lo mismo que acceder a ellos

La forma más sencilla de gestionar sets de datos es usando Google Drive, la herramienta que reemplazó a Google Docs por sus mayores y mejores funcionalidades. Es necesaria su libre descarga a partir de una cuenta de Gmail y, una vez instalada, permite alojar todo tipo de documentos, en el caso que nos ocupa, desde hojas de cálculo en Excel hasta tablas, gracias a su novedosa función Google Tablas Dinámicas (<https://support.google.com/fusiontables/answer/2571232?hl=es>).

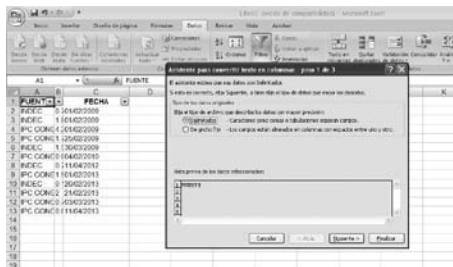
Las formas más comunes de gestionar y alojar sets de datos se llevan a cabo mediante planillas de Excel o mediante un formato denominado CSV, igualmente abierto y reutilizable, en el que las columnas están separadas por comas.

Cuando se descarga un fichero de datos en CSV, la imagen que se obtiene es del siguiente tipo:

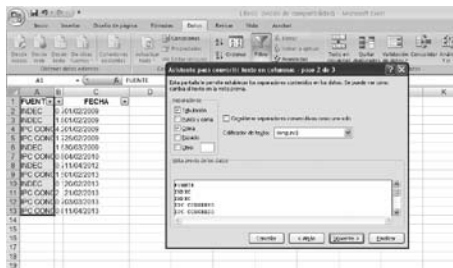
FUENTE	FECHA
2. INDEC	01/01/2009
3. INDEC	11/01/2009
4. IPC CONGRESO	01/01/2009
5. IPC CONGRESO	15/02/2009
6. INDEC	11/03/2009
7. IPC CONGRESO	04/03/2010
8. INDEC	01/10/2012
9. IPC CONGRESO	15/01/2013
10. INDEC	01/30/2013
11. IPC CONGRESO	21/02/2013
12. IPC CONGRESO	03/03/2013
13. IPC CONGRESO	11/04/2013

En estos casos, hay que volver a estructurar las columnas. El procedimiento es sencillo, sin mover el cursor de la Fila A:

- 1.- Ir a la pestaña “Datos”
- 2.- Hacer clic en “Texto en columnas”
- 3.- Se abrirá un cuadro de diálogo, marcar “Delimitados” + “Siguiete”



- 4.- Marcar “Tabulación” + “Comas”



5.- Nótese que, durante el procedimiento, la columna A permanece marcada o “pin-tada”. Luego, hacer clic en “Siguiete” y lo que se obtendrá será este resultado:

FUENTE	VARIACION	FECHA
2. INDEC	0,7	01/01/2009
3. INDEC	1,87	01/01/2009
4. IPC CONGRESO	4,2	01/01/2009
5. IPC CONGRESO	1,7	29/01/2009
6. INDEC	1,5	30/01/2009
7. IPC CONGRESO	0,8	04/03/2010
8. INDEC	0,7	11/03/2010
9. IPC CONGRESO	1,52	01/02/2013
10. INDEC	0,11	30/01/2013
11. IPC CONGRESO	2	21/02/2013
12. IPC CONGRESO	0,7	03/03/2013
13. IPC CONGRESO	0,8	11/04/2013

6.- Para lograr el resultado que ofrece la vista de arriba se centraron los valores de las celdas y solo resta colocar las variables que no aparecen en las columnas B y C, que originalmente eran “Variación” y “Fecha”.

FUENTE	VARIACION	FECHA
INDEC	0,7	1/31/2011
INDEC	1,87	1/31/2011
IPC CONGRESO	4,2	2/28/2011
IPC CONGRESO	1,7	2/28/2011
INDEC	1,5	3/31/2011
IPC CONGRESO	0,8	3/31/2011
INDEC	0,7	4/30/2011
IPC CONGRESO	1,52	4/30/2011
INDEC	0,11	5/31/2011
IPC CONGRESO	2	7/30/2011
IPC CONGRESO	0,7	6/30/2011
IPC CONGRESO	0,8	6/30/2011
INDEC	1,87	7/30/2011
IPC CONGRESO	0,8	8/31/2011
INDEC	1,89	8/31/2011

Con este último paso, el procedimiento se completa. El lector puede probar el procedimiento descargando este set de datos en formato CSV en: <http://data.lanacion.com.ar/datastreams/69218/inflacion-indec-y-estimaciones-privadas-variacion-mensual/>. En este caso, se trata de un set

perteneciente al catálogo de datos abiertos del diario *La Nación* de Argentina.

Estos sets están alojados dentro de una plataforma denominada Junar. Con una cuenta gratuita, este recurso permite gestionar hasta cinco sets de datos. La plataforma posibilita a las empresas, Gobiernos y otras organizaciones liberar sus datos y exponerlos como datos abiertos, impulsando la colaboración y la transparencia.

Crónicas cortas, con conclusiones contundentes y una gran visualización

La Plataforma Junar Open Data permite recoger, refinar y publicar al mismo tiempo. Varios Gobiernos, medios de comunicación como *La Nación* de Argentina y ONG la usan para compartir sus colecciones de datos con la audiencia. La pestaña inferior izquierda de cada set de datos, denominada “Acciones”, permite la exportación de información tanto en Excel como en CSV, así como obtener un código para incrustar en un blog o sitio web, visualizar como Google Hojas de Cálculo y socializar a través de Twitter, Facebook, LinkedIn y Google Plus.

Hay otras plataformas desde donde se pueden alojar, gestionar y descargar sets de datos, como Socrata, de la que se valen igualmente tanto medios de comunicación como varias ONG. Un ejemplo es la colección de set de datos denominada Finances, del Banco Mundial. Todos

los sets de datos financieros del Banco Mundial están aquí: <https://finances.worldbank.org/page/datasets>. El ejemplo más conocido dentro del campo periodístico es el que se visualiza en este enlace: <https://finances.worldbank.org/Procurement/Major-Contract-Awards-FY2007-FY2013-Beta-version/kdwi-wcs3>, que contiene información completa de todos los préstamos otorgados desde 2007 hasta la actualidad por países.

A la derecha de cada columna, hay una flecha que abre un cuadro desde donde se pueden seleccionar opciones como ocultar columnas o seleccionar un solo país.

Desde el momento en que se activa el proceso de selección de datos, la plataforma permite guardar esa selección como propia, descargarla en el PC, así como visualizar los datos de la serie, entre otras funciones.

Para ver más recursos en Socrata, se puede consultar en: <https://opendata.socrata.com/>.

Visualización para principiantes

El periodismo de datos maneja un volumen grande de datos. Sin una adecuada visualización que acompañe el análisis de esos datos, sería imposible visualizar sencillamente las variables que se ponen en estudio. Las crónicas deberían ser extensas y podrían pecar de aburridas si están plagadas de números que la gente no entienda con facilidad al no ser expuestos adecuadamente. Por ello, se sugieren crónicas cortas, con conclusiones contundentes y una visualización

importante, dominante en el artículo, a la hora de contar una historia.

Los expertos en visualización interactiva son capaces de diseñar, a partir de una gran cantidad de datos, un gráfico que permite al lector decidir qué desea visualizar y cómo. La visualización de datos explica de manera comprensible las relaciones entre gran cantidad de información que se genera por distintas vías.

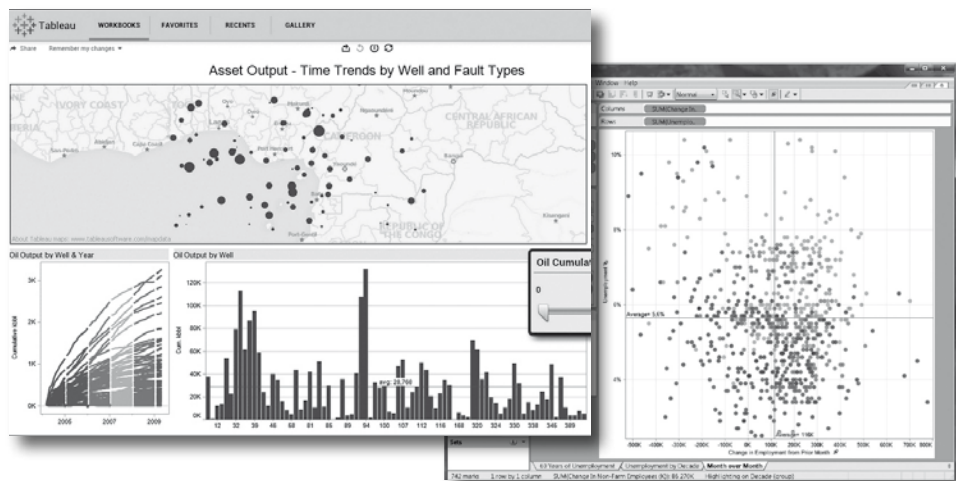
Para principiantes en visualización: ChartGo y, después, Many Eyes

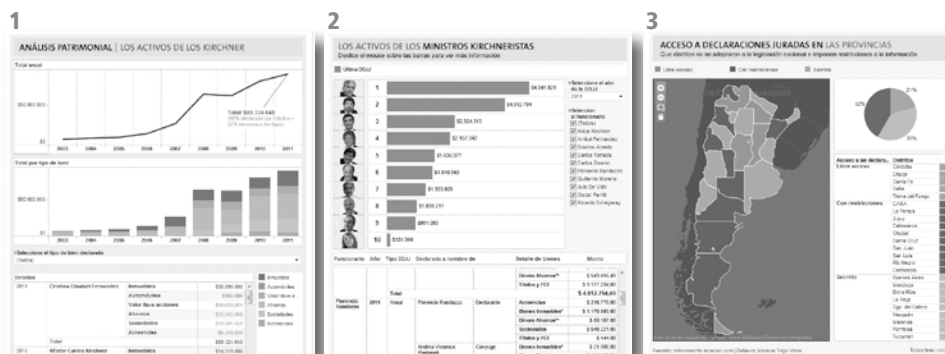
El clásico ejemplo para el reportero es el presupuesto público. El Ministerio de Economía entrega un pesado reporte lleno de tablas y números que no producen ninguna información interesante a simple vista ni, mucho menos, una noticia

que no sea la anunciada en el parte oficial de prensa. Aunque un presupuesto contiene variables, es decir, elementos que son medidos a lo largo del tiempo. Y si se dispone de un presupuesto del año en curso, probablemente tenga a mano el del año anterior, que tiene... las mismas variables. Sin embargo, muchas veces, hasta la más sencilla de las tablas puede resultar muy aburrida.

La comparación es quizá una de las prácticas más usadas por el periodismo, pero, cuando hay que procesar gran cantidad de información numérica, el asunto se complica. Por ello, se recurre a distintas formas de visualizar el conjunto de datos; de ese modo, podemos analizar mejor ciertos procesos y cambios no detectados de un vistazo.

Uno de los muchos ejemplos que podemos mencionar es la herramienta Tableau Public:





Hay una larga lista de ejemplos que el diario *La Nación* de Argentina puede mostrar en el uso de este recurso. Por ejemplo, las siguientes:

Declaraciones juradas patrimoniales¹

- Tipo de representación visual: fiebre + barras + tabla / barras + tabla
- Interactividad: menú desplegable + *tooltip* con datos al *roll over*
- Programa: ambos en Tableau Public

Declaraciones juradas patrimoniales²

- Tipo de representación visual: barras + imágenes fotográficas + tabla / círculos apilados + tabla
- Interactividad: menú desplegable, selector + *tooltip* con datos al *roll over*
- Programa: ambos en Tableau Public
- Extras: enlaces a Documents Cloud en el *tooltip*

Declaraciones juradas patrimoniales³

- Tipo de representación visual: geolocalización con polígonos + torta + tabla
- Interactividad: *tooltip* con datos y filtro al *roll over*
- Programa: Tableau Public

Desempleo en Argentina⁴

- Tipo de representación visual: fiebres + tabla
- Interactividad: menú desplegable + selector + *tooltip* con datos al *roll over*
- Programa: Tableau Public

Subsidios a transporte público de pasajeros⁵

- Tipo de representación visual: semicírculos y barras apiladas
- Interactividad: filtro al *roll over* y aparición de barras al clic
- Programa: Flash

- 1.- <http://www.lanacion.com.ar/1525583-el-crecimiento-de-los-bienes-de-los-kirchner-de-7-a-89-millones-de-pesos>.
- 2.- <http://www.lanacion.com.ar/1535554-la-evolucion-de-la-fortuna-personal-de-los-ministros-kirchneristas>.
- 3.- <http://www.lanacion.com.ar/1546537-que-provincias-esconden-bajo-llave-el-patrimonio-de-sus-gobernadores>.
Enlace directo: <http://public.tableausoftware.com/shared/Z47YDQC9M>.
- 4.- <http://www.lanacion.com.ar/1552132-el-desempleo-en-personas-como-vo>.
Enlace directo: <http://public.tableausoftware.com/shared/TWDPMN4P>.
- 5.- <http://www.lanacion.com.ar/1380725-colectivos-insaciables-un-cheque-diario-de-10-millones-en-subsidios>.

Asimismo, hay hoy otros recursos que compiten por la delantera, como Google Fusion Tables, que permite diseñar desde cero una tabla y trasladar esa información hacia un mapa. Hay una galería de imágenes que se puede consultar en: <https://sites.google.com/site/fusiontablestalks/stories>. Por supuesto, se puede comenzar desde cero, aunque eso requiere entrenamiento. Se puede ver un tutorial en: <http://support.google.com/fusiontables/answer/184641?hl=en>. Y más en: <http://pushmatrix.github.com/fusiontable-slides/#/2>.

A modo ilustrativo, el proyecto Censo del *New York Times* ha utilizado este recurso para mostrar los datos del censo 2010: <http://projects.nytimes.com/census/2010/map>. Pero no es el único, ya que otros medios han usado esta herramienta para el mapeo interactivo del censo, como el *Texas Tribune* (<http://www.texastribune.org/library/data/census-2010/>) y el *Washington Post* (<http://www.washingtonpost.com/wp-srv/special/nation/census/2010/>).

Sin embargo, estos recursos no son sencillos de usar para un principiante. Requieren entrenamiento y mucha práctica, aunque esto no debe ser visto como un impedimento, sino como un desafío. Un ejemplo es el del bloguero Andy Tow, quien usa la herramienta con frecuencia para su blog de datos, tal como se visualiza aquí: <http://towsa.com/wordpress/>. Entonces, la conclusión es: si Andy puede hacerlo, otros pueden seguir su ejemplo.

No obstante, hay recursos sencillos que posibilitan graficar de manera interactiva

sin necesidad de mayores conocimientos. Quizá el más usado y versátil es Many Eyes: permite diseñar varios tipos de visualizaciones diferentes, creando una cuenta gratuita; no es necesario descargar ningún programa (lo cual es ventajoso para muchos periodistas que trabajan en redacciones en las que, por cuestiones de seguridad informática, sus terminales no les permiten la bajada de programas).

Sugiero a los principiantes que comiencen con ChartGo (<http://www.chartgo.com/>), el cual usé experimentalmente y sin experiencia previa alguna en 2010, cuando comenzaba a incursionar en periodismo de datos (ver nota en: <http://www.sololocal.info/noticias/1-de-bahia/1968-pauta-oficial-el-reparto-de-la-torta-.html>).

Luego, un segundo paso sería aprender el uso de Many Eyes, al que considero muy conveniente por las razones antes apuntadas.

Esto va dirigido especialmente al reportero que no dispone de un profesional especialista en la materia dentro de su equipo: esta es la realidad de miles de reporteros, blogueros y periodistas de portales de noticias en línea, que deben cumplir varias funciones al mismo tiempo por carencia de recursos.

Afirmar que para hacer periodismo de datos se requiere de un equipo visualizador especializado como condición sine qua non sería establecer una brecha demasiado grande y peligrosa entre unos y otros.

Por eso, los recursos más complicados de usar quedan en manos de pro-

fesionales o en las de aquellos que se lanzan a descubrir como autodidactas la maravillosa tarea de llevar a una imagen algo tan aburrido como una tabla. Y para quienes no hemos recibido instrucción formal en estas prácticas nos quedan los recursos más sencillos o los que yo llamo “*ready to use*”, es decir, listos para usar.

Gráficos interactivos inteligentes

La visualización de datos está teniendo en nuestros días un desarrollo espectacular, que impacta directamente en el periodismo; no solamente en medios digitales, también la prensa gráfica y la televisión se están valiendo de estas herramientas para presentar visualmente sus historias de una forma más interesante.

Hace algunos años, para ser periodista, con buscar datos, abordar fuentes y saber escribir era suficiente. En nuestro tiempo, eso ha cambiado, porque los periodistas comenzamos a percibir que, si no manejamos adecuadamente estos recursos, podemos incluso perder la oportunidad de encontrar primicias. Esto requiere que el reportero desarrolle habilidades especiales.

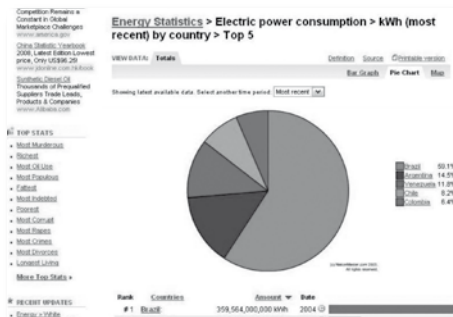
Buena parte de la información se encuentra en bases de datos u oculta en las webs. A finales de los 70, cuando comencé mi carrera como reportera, el escenario era de pocos datos a la vista, por lo que los periodistas de mi generación debimos adaptarnos a trabajar en un terreno más dominado por las fuentes orales que por las documentales.

Eso ha cambiado en el siglo XXI, en el que la sobreabundancia de información puede incluso sumirnos en un caos. Antes, el reportero escribía y el equipo de infografía hacía su labor sobre la base de los datos que el periodista le proporcionaba. Esa relación ha cambiado drásticamente en los últimos años.

Ahora, para encontrar noticias, el periodista necesita muchas veces apoyarse en herramientas visuales como Google Maps y Google Earth u otras similares como la de generación de gráficos.

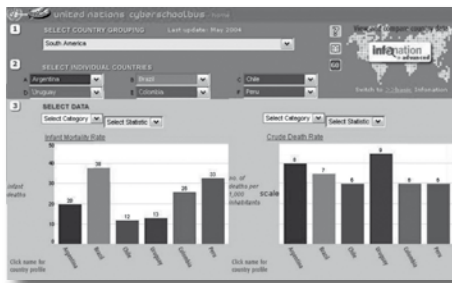
Por ejemplo, varias crónicas periodísticas de los años 90 fueron ilustradas o surgieron sobre la base de recursos tales como NationMaster (www.nationmaster.com), que permite obtener gráficos e información comparada entre varios países del mundo.

Se ingresa mediante la pestaña “Countries A-Z”, desde donde se seleccionarán las variables. La imagen inferior (<http://www.nationmaster.com/countries>) corresponde a una captura de pantalla de un gráfico de torta realizado a través de NationMaster para comparar el consumo de energía en cinco países de América del Sur.



Otro similar, muy usado en el pasado, era InfoNacion, un sitio web educativo de las Naciones Unidas de uso frecuente en liceos y universidades, al que se accede mediante el siguiente enlace: http://cyberschoolbus.un.org/infonation/s_infonation.htm.

La imagen de abajo (<http://cyberschoolbus.un.org/infonation3/ie&nn4.html>) corresponde a una captura de pantalla del recurso para la comparación de distintas variables correspondientes a seis países de América del Sur. Generar las gráficas demandó apenas 40 segundos.



Estos recursos, sin embargo, poseen la desventaja de que los datos no están tan actualizados como quisiéramos. Por ejemplo, varios indicadores corresponden al año 2004. Ya nadie los usa, salvo como referenciales.

Los administradores de medios independientes o blogueros han debido aprender a utilizar otros de más reciente aparición o que ofrecen mayores ventajas.

Tablas en línea

Insertar una hoja de cálculo en una página web puede resultar un problema si no se tiene mucha experiencia con Excel

o si lo que se desea es obtener el código HTML que traduzca esa información de un modo sencillo y amigable.

COUNTRY	POPULATION	RANK
China	1,34,73,50,000	1
India	1,21,01,93,422	2
USA	31,48,90,000	3

Tableizer (<http://tableizer.journalistopia.com/>) convierte cualquier hoja de cálculo en una tabla HTML. Lo único que hay que hacer es copiar el contenido de la hoja de cálculo y pegar esa información en el cuadro que tenemos a nuestra disposición. Posteriormente, se pueden configurar preferencias en cuanto al formato y, automáticamente, un botón genera el código. El resultado es una vista muy sencilla pero muy útil, ya que en menos de unos minutos se puede obtener el resultado. En la imagen de arriba, en la parte izquierda, se puede observar la introducción de datos en Tableizer; a la derecha, lo que resulta de pegar el código HTML (fuente: <http://img.labnol.org/di/tableizer2.gif>).

Sistemas de información geográfica

Se los define como sistemas de integración de datos, capaces de mostrar información geográficamente localizada. Los datos censales, por ejemplo, suelen ser aburridos; aunque, administrados mediante programas de SIG, pueden ofrecer una visión clara de las variables en estudio.

Los usuarios de estos sistemas pueden crear mapas interactivos, poner bajo la lupa información espacial y editar datos y mapas, de tal modo que se ofrezcan mejores representaciones de la realidad.

En palabras sencillas, un SIG es un *software* que permite una mejor representación visual de datos numéricos, conforme estos se distribuyan geográfica y espacialmente.

El Instituto de Periodismo Analítico (<http://www.analyticjournalism.com/>) viene trabajando desde hace años en la incorporación de estas herramientas al campo de las comunicaciones. El sitio contiene una gran cantidad de valiosas herramientas que serán de mucha ayuda para docentes de periodismo y reporteros interesados en el tema.

Un SIG permite una mejor representación visual de datos numéricos

Para comprobar cómo los sistemas de información geográfica impactan en el periodismo, se puede ver la ponencia de Tom Johnson, director del Instituto de Periodismo Analítico, en el siguiente enlace: <http://www.slideshare.net/jtjohnson>.

Las universidades y empresas privadas en América Latina vienen interesándose por esta tecnología desde hace algunos años e, igualmente, los Gobiernos incorporan los SIG a distintas áreas de gestión.

También en el área de transporte, el recurso <http://mtcgeo.mtc.gob.pe/website/telecom/viewer.htm> es de libre consulta, como el relacionado con vivienda, por lo que es fácil imaginar que estas herramientas dan soporte esencial para la visualización de datos en estas materias.

En Argentina, por ejemplo, el sitio <http://www.mapaeducativo.edu.ar/> ofrece una introducción a estos sistemas, tanto para docentes como para alumnos; pero también hay herramientas para periodistas, como la localización de escuelas y su distribución espacial, que bien podrían formar parte de una crónica.

Por su parte, el sitio <http://www.sig.gov.ar/> reúne información de diferentes oficinas estatales, generando recursos de libre uso para los usuarios registrados.

Los sistemas de información geográfica disponibles en una larga lista de plataformas de E-Gov mejoran el acceso a la información pública por parte de los ciudadanos, por lo que los periodistas no deberían quedar al margen de este proceso. En el futuro, consultar estos sistemas será tan frecuente como hoy en día lo es consultar el pronóstico del tiempo en cualquier portal sobre meteorología.

Así como un reportero puede consultar estos sistemas, también puede manejarlos, cosa que requiere un cierto entrenamiento y práctica. La mayoría de los programas de SIG son de pago, como el *software* ArcView (<http://www.esri.com/>), que sigue siendo

uno de los SIG estándar más utilizados en el mundo. Si bien la mayoría de los medios está usando ArcView, otros utilizan Maptitude, ya que tiene una interfaz en español, como se ve en el enlace <http://www.caliper.com/maptitude/MappingSoftware.htm>, así como la versión profesional de MapInfo, que se obtiene en el sitio <http://www.pbinsight.com/>, la cual dispone asimismo de una versión básica gratis, aunque en inglés.

Gráficos listos para usar

- **Sistema Regional de Visualización y Monitoreo (Servir):** una serie de recursos interactivos se encuentra disponible en <http://www.servir.net/america-latina-caribe>, que integra observaciones de la tierra con modelos de pronóstico y datos del escenario geográfico. En Centroamérica y el Caribe, el equipo de Servir ha dado respuesta a más de 20 desastres naturales y diez amenazas ambientales. El portal geoespacial permite el acceso a los datos y metadatos de la región.

- **Portal de datos:** <http://maps.cathalac.org/Portal/>.

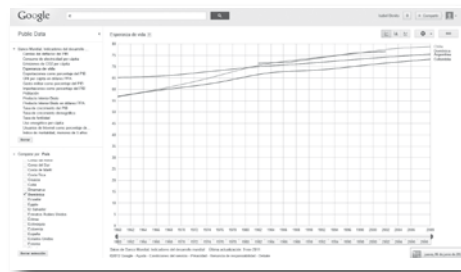
- **Visualizador de mapas:** <http://share1.cathalac.org/servirmaps/index1.html>.

- **Pronóstico para Mesoamérica:**
http://www.servir.net/MM5_Mesoam
%C3%A9rica.

- **Análisis de desastres:** http://www.servir.net/An%C3%A1lisis_de_Desastres/Desastres/.

- **GapMinder:** muchos de nosotros conocemos <http://www.gapminder.org/>, un clásico en este tema. Explorar esta herramienta puede resultar un excelente ejercicio, a fin de comprender sus potencialidades. Para obtener gráficas o mapas en GapMinder, el acceso (a la izquierda) es GapMinder World, en el que en la barra inferior se puede seleccionar la variable y, en el cuadro de la derecha, marcar el país o los países que se desean incluir. La página posee un tutorial en vídeo y en PDF.

- **Google PublicData:** http://www.google.com/publicdata?ds=wb-wdi&met=sp_dyn_le00_in&idim=country. Una herramienta interesante que permite analizar la expectativa de vida entre países de un modo muy sencillo: simplemente, se señala los países que se desea comparar y se obtiene un resultado como el que se muestra para Chile, Argentina, Colombia y República Dominicana en la captura de pantalla:



El servicio es muy sencillo de usar, recomendado especialmente para quienes no están muy familiarizados con el manejo de gráficos. Posee otras herramientas como Data Visualizer (<http://devdata.worldbank.org/DataVisualizer/>), que permite realizar gráficas sobre cualquiera de los 49 indicadores para 209 países. Las variables pertenecen a los campos de la economía, finanzas, información general, tecnología y medio ambiente.

Los *mashups* permiten combinar datos de diferentes fuentes

- **Google Finance:** <http://www.google.com/finance>. Para quienes cubren la sección de economía y finanzas, Google Finance es otra herramienta eficaz. En especial, si se trata de obtener gráficos sobre compañías públicas y privadas, ya que organiza la información relacionada con el mundo financiero: cotizaciones en Bolsa y datos financieros de las empresas. La búsqueda por nombre de empresa, como la que se ve en la captura de pantalla para Dow Chemical, una compañía que opera en la ciudad de Bahía Blanca (Argentina), proporciona datos en tiempo real, lo cual constituye un notable aporte para quienes trabajan en medios que manejan información de último momento.

Esta herramienta tiene un antecesor, Yahoo Finance: <http://espanol.finance>.



yahoo.com/. Esta se ha convertido en la favorita de muchos periodistas, ya que ofrece las cotizaciones de la Bolsa, tasas de cambio de valores, comunicados de prensa de empresas, además de algunas herramientas de gestión para finanzas personales. Pero habrá que probar uno y otro servicio para escoger el que se adapte mejor a las preferencias de cada usuario.

- **Google Maps:** <http://maps.google.com/>. Divide el mapa completo en pequeñas imágenes con un tamaño de 256 x 256 píxeles. Permite encontrar y mostrar rutas y direcciones para llegar a un lugar determinado. Una interesante característica es la de brindar distintas vistas de los mapas: la satelital y la que corresponde al relieve. El cursor se arrastra permitiendo los movimientos hacia cualquier dirección, mientras que la herramienta de zoom y alejamiento nos da una vista puntual o más general del espacio en estudio.

Las imágenes de Google Maps son tomadas mediante un satélite de teledetección, que viene funcionando desde 2001, ubicado a menos de 500 kilómetros sobre la superficie terrestre, en el espacio.

Aunque aún hay zonas geográficas del globo terráqueo que no tienen acce-

so a las vistas, actualmente constituye una herramienta de uso cotidiana para los reporteros.

Cualquier usuario con una cuenta de Google puede ingresar al sistema, buscar sus propios mapas, subir fotos y, de esta manera, ubicarse geográficamente en el mundo o ubicar su comercio, oficina o empresa, siendo útil para el mercadeo de actividades comerciales u organizacionales.

Las aplicaciones se multiplican: si se instala Google Gears (<http://gears.google.com/>) en nuestro navegador, va a aparecer un botón en la interfaz de Google Maps sobre la herramienta de zoom y bajo la herramienta de vista panorámica que, al pulsarlo, nos mostrará la posición en la que nos encontramos. Esta ubicación se obtiene a través de las redes wifi disponibles o mediante nuestra IP.

Otro recurso interesante que aparece en el menú es el llamado ¿Qué hay aquí?, con el que se puede encontrar el resultado que mejor represente a ese lugar, ya sea una dirección específica, parte del paisaje o el nombre de un sitio. Esta función tiene en cuenta el nivel de zoom que se está mirando y ofrece el dato geográfico más apropiado desde ese punto de vista.

- **Google Earth:** <http://earth.google.es/>. Como el propio nombre del recurso dice, Google Earth te permite volar a cualquier lugar de la Tierra y ver imágenes de satélite, mapas, relieve y hasta edificios en 3D. Ahora también puedes ver galaxias

del espacio exterior y hasta sumergirte en el océano. Se puede explorar un amplio contenido geográfico, guardar los lugares que se visitan y compartirlos con otras personas. El programa es de libre descarga y se accede en el siguiente enlace: <http://earth.google.es/download-earth.html>. Las imágenes que muestra Google Earth son obtenidas mediante un satélite llamado QuickBird, lanzado al espacio en octubre de 2001.

El recurso es muy útil cuando se investigan compañías o casos de propiedades adquiridas por funcionarios. O, por ejemplo, el impacto ecológico de la polución.

Puede citarse como caso práctico la crónica “Discurso oficial: a contramano de Google Earth” (<http://tinyurl.com/acntramanodegoogleeahrt>), que representa la manera más sencilla de presentar esta utilidad en el contexto de una noticia. Hay, por supuesto, otras mucho más sofisticadas.

Un dato que se debe tener en cuenta es que la búsqueda avanzada de Google permite encontrar lugares, personas, empresas o grupos seleccionando el formato que devuelve imágenes registradas por Google Earth. Esto resulta de mucha utilidad; en especial, cuando hay que monitorear movimientos portuarios, tal como se observa en la imagen disponible en <http://www.portalwhite.com.ar/archivos/IngenieroWhite-BuenosAires-Argentina.kmz>, obtenida tras una búsqueda en la que se introdujo en el cuadro para “la frase exacta” el término “Ingeniero Whi-

te” –que corresponde al puerto que funciona en Bahía Blanca– y se seleccionó el formato KMZ, que es el que devuelve vistas de Google Earth. El otro formato apto para selección es el KML.

Los formatos KMZ y KML sirven para representar datos geográficos. Un fichero KML especifica un lugar, una imagen o un polígono. Puede contener un título, una descripción básica del lugar, sus coordenadas (latitud y longitud), así como otros datos. En cuanto al formato KMZ, no es más que un fichero KML comprimido. En concreto, KMZ usa el formato de compresión ZIP.

Aplicaciones híbridas: ‘mashups’

Existe una variada gama de aplicaciones de Google Maps y de Google Earth. Por ejemplo, Eosnap (<http://www.eosnap.com/>) es un portal dedicado a la observación de la Tierra desde el espacio. Cada día, el sitio genera imágenes satelitales que resultan de indudable interés en el campo de las comunicaciones.

Muchos de estos recursos son híbridos (*mashup* o *remezcla*), es decir, sitios que usan contenidos de otras aplicaciones para crear nuevas funciones. Los *mashups* permiten la combinación de datos que existen en diferentes fuentes, como YouTube, Flickr, Google Maps y Google Earth, entre otras muchas.

El escenario en el que se desarrollan los *mashups* es increíblemente versátil y todos los días aparecen nuevos recursos. A estas alturas, cualquier reportero podría sentir algo de fastidio por el hecho

de que la mayoría de estas herramientas esté disponible en inglés. Ciertamente, el reportero que no domina ese idioma –para su lectura, al menos– tendrá dificultades para ponerse al día en el uso de estas tecnologías. No obstante, se están desarrollando aplicaciones híbridas en español, que tienen un enorme potencial. Entre esos *mashups* destacan:

- **Tagzania** (<http://tagzania.com/>): permite etiquetar lugares, por lo cual los usuarios forman una base de datos colaborativa de enorme potencial. Cualquier periodista podría sorprenderse cuando introduce el nombre de la ciudad donde vive, ya que tal vez encuentre más resultados de los que espera, como ocurrió en mi caso. Un recurso similar es Flof (<http://flof.com.ar/bin/home/>).

- **WikiLoc** (<http://es.wikiloc.com/wikiloc/home.do>): para marcar rutas, senderos, trayectorias de caminatas o ciclismo. Combina el uso del GPS, cuando se crean puntos de interés.

- **Woices** (<http://woices.com/>): encontrar o grabar audios geolocalizados.

- **Proyecto Simile** (<http://simile-widgets.org/>): es una cita obligada dentro del campo de la visualización de datos. Semantic Interoperability of Metadata and Information in Unlike Environments (Simile) es un proyecto de investigación que apunta a maximizar la interacción entre diferentes colecciones digitales. Buena

parte de la información contenida en bibliotecas digitales resulta invisible para los motores de búsqueda en línea. Por eso, este recurso agrupa una interesante variedad de herramientas que tratan de resolver ese inconveniente. Simile es un recurso desarrollado por The World Wide Web Consortium (www.w3c.org), con el apoyo del Instituto Tecnológico de Massachusetts (EE. UU.), entre otras organizaciones que participan en la iniciativa.

Las herramientas desarrolladas en Simile utilizan las tecnologías de web semántica para mejorar el intercambio de los recursos web, lo que ayuda a desbloquear el contenido de las bibliotecas digitales de todo el mundo.

Su recurso más conocido es Piggy Bank (http://simile.mit.edu/wiki/Piggy_Bank), una extensión de Firefox que permite introducir al navegador en la web semántica, facilitando el procesamiento de páginas webs, de tal modo que se logre recuperar información no visible desde la web. Una vez instalado, aparecerán dos iconos, que deberán añadirse a la barra de herramientas.

Pero no es el único. Exhibit (<http://www.simile-widgets.org/exhibit/>), por ejemplo, posibilita la creación de presentaciones de elementos mediante un carrusel, que puede ser personalizado gracias a la selección de varios parámetros.

Otra maravilla de este proyecto es Time Plot (<http://www.simile-widgets.org/timeplot/>), sistema que permite graficar estadísticas de manera original, dando la opción de insertar una línea

temporal mediante la incorporación de eventos concretos.

Por su parte, Citeline (<http://citeline.mit.edu/>) facilita la edición de bibliografías interactivas.

Condiciones necesarias

La respuesta a cuáles son las condiciones mínimas necesarias para procesar datos desde una sala de redacción y llegar a conclusiones útiles a los ciudadanos es compleja. Vivimos en un mundo en el que casi todo se expresa con números. Por lo tanto, el manejo de grandes masas de datos requiere ciertas habilidades por parte de quienes pretenden encontrar noticias relevantes detrás de ellos.

La misión no es sencilla: se tiene que encontrar un significado tangible a partir de lo abstracto, hay que contarle a la gente cómo esa gran maraña de números afecta su vida. Solo de ese modo se logra el impacto necesario para vivir en sociedades mejor informadas.

El periodismo de datos requiere entrenamiento permanente y conocimiento de materias generalmente alejadas de las preferencias del periodista convencional. Las fuentes aceitadas no son suficientes: hay que conocer el funcionamiento de la Administración Pública, saber interpretar leyes, decretos y ordenanzas. Si un periodista no comprende cómo funcionan las instituciones de su país, difícilmente podrá emprender el camino con éxito.

- Dominio del idioma inglés: la mayoría de las herramientas de uso frecuente en periodismo de datos no tiene corre-

lato al español. Actualmente, se está formando una brecha importante entre quienes acceden a estos recursos por su condición bilingüe y los que no.

- Habilidades desarrolladas en materia de búsquedas en la web: con usar Google de manera tradicional no alcanza. Hay que aprender a sacar mayor provecho de las búsquedas avanzadas, por formato, por dominio y por fecha; pero también, ante la carencia de una iniciativa nacional de datos abiertos, controlar los posibles sitios oficiales en los que hay datos.

El periodismo de datos requiere entrenamiento permanente

- Conocer los principios básicos del programa Excel o similar, capaz de procesar información numérica en hojas de cálculo.

- Tener conocimientos de matemática básica y de estadística, tanto descriptiva como inferencial: un periodista “anumérico” no tendrá la capacidad suficiente para enfrentar el proceso conocido como “la entrevista a los números”. Esto nos interesa especialmente cuando se trata de análisis de presupuestos públicos, impuestos y otros temas como criminalidad y contaminación ambiental.

El periodista se encuentra frente a un escenario desconocido. Hemos sido formados para un mundo en el que prevalece la carencia de datos, no para un mundo en el que hay sobreabundancia de

información. Los datos nos llegan desde varios frentes y en grandes cantidades.

No hay todavía instrucción formal sobre periodismo de datos en las universidades, salvo excepciones. Por lo tanto, los futuros periodistas de datos deberán invertir –básicamente, en tiempo– para afrontar su propia capacitación en esta disciplina.

El periodismo de datos implica un trabajo duro y, muchas veces, costoso. Hay que acceder a decenas, cuando no a cientos o a miles de planillas, textos, gráficos. Hay que recurrir, en ocasiones, a programadores para que diseñen aplicaciones capaces de capturar datos contenidos en páginas webs. El esfuerzo siempre será un componente con fuerte presencia en esta materia. A veces, la rutina es decepcionante, porque no siempre encontrará lo que busca. Tendrá que leer pesados expedientes, montañas de papeles, hacer llamadas que nadie le devolverá y llamar a puertas que probablemente no se abrirán. El periodista de datos siempre deberá considerar la posibilidad de que, en ocasiones, no hay noticias detrás de los datos. Aunque claro, eso no es lo más habitual.

Ahora que sabe lo peor del periodismo de datos, si desea seguir adelante, entonces está preparado para enfrentar el desafío; y, seguramente, lo logrará, porque en esta bendita profesión no brillan las personas con suerte, con buenas conexiones, ni siquiera con mentes brillantes: en esta tarea solo brillan los perseverantes. Ahí está la diferencia. ■