

EpData: informando sobre el coronavirus

Iñaki Hernández

*Coordinador Sección de internet y EpData.
Europa Press*

La última vez que lo comprobamos, llevábamos más de doce versiones del script de R que usamos en EpData para ‘escrapear’ el pdf del Ministerio de Sanidad con los datos del balance del coronavirus en España. Cada cambio de criterio de Sanidad, cada dato que decidíamos incorporar a nuestro seguimiento, nos obligaba a retocar el código y añadir una nueva versión hasta superar la docena. Y eso fue antes de que el Gobierno decidiera dar un giro, una vez más, a su estrategia de seguimiento y anunciara que dejaba de medir la incidencia para la población general. Ahora ya, hemos perdido la cuenta.

Es una anécdota, pero de algún modo ejemplifica cómo esta pandemia nos ha obligado a todos a adaptarnos e ir aprendiendo sobre la marcha. Cuando decidimos cubrir desde EpData el nuevo brote de coronavirus, se trataba de una noticia limitada a China. De hecho, la página web que lanzamos para seguir la enfermedad se llamaba ‘Coronavirus en China, en gráficos’.

Durante semanas, fuimos testigos de cómo los datos de China iban empeorando. Con dificultades por la poca transparencia del país, lo fuimos reflejando en gráficos. Luego llegó Italia. Y de repente nos encontramos encerramos en casa, con un proyecto que se llamaba coronavirus en China, pero que, de una manera inimaginable, al menos inimaginable hace menos de un mes, nos afectaba directamente.

Fueron días complicados, en los que la información llegaba “a cuentagotas”. La primera versión del balance de Sanidad, de febrero de 2020 –ahora ya vamos por unos 600 balances– daba información sobre la incidencia, nuevos casos y fallecidos, pero no ofrecía información desagregada ni una serie en la que ver la evolución ni datos en formato reutilizable.

Rápidamente vimos la necesidad de ofrecer a los medios y usuarios información de lo que estaba pasando en un formato abierto. Creamos series de datos que actualizamos diariamente con ese balance. Según la situación empeoraba, y los balances de Sanidad ganaban en complejidad, nosotros también íbamos sofisticando nuestra oferta de datos en abierto.

La gente buscaba desesperadamente información sobre la enfermedad. El tráfico de la web se disparó con búsquedas de personas anónimas, periodistas e investigadores que usaban las series que íbamos construyendo día a día para sus reportes. Medios locales y nacionales comenzaron a usar nuestros gráficos para insertarlos en las noticias sobre la pandemia.

La falta de información, al principio, y los cambios de criterio obligados por la situación después, nos forzaron a nosotros también a ir tomando decisiones difíciles. Cuando se empezó a debatir sobre la necesidad de mirar a la cifra de fallecidos totales, ante la limitación que suponía el número de fallecidos confirmados de COVID-19 por el colapso del sistema, recurrimos a los registros civiles para construir una nueva serie. Cuando se demostró la buena correlación que existía entre los desplazamientos y los repuntes de casos, incorporamos esa información. Aprendíamos y crecíamos sobre la marcha.

No lo tuvimos que construir desde cero. EpData nació en 2017 como la apuesta de Europa Press por el periodismo de datos. Financiado por la iniciativa Digital News Initiative (DNI, por sus siglas en inglés) de Google, la idea de EpData es trasladar al mundo del periodismo de datos la filosofía de una agencia de noticias.

Mucho antes de que se hablara de periodismo de datos en España, los periodistas de la agencia de noticias ya trabajaban con estadísticas. Informaban del paro, de la evolución de casos de violencia de género o de la incidencia de la gripe, por poner tres ejemplos dispares, y contextualizaban el dato puntual, noticioso, con sus propias series estadísticas, recopiladas

durante años y guardadas en ocasiones en libretas o cuadernos en cajones.

Lo que hicimos al crear EpData fue volcar todo ese conocimiento que ya existía en la agencia en una gran base de datos compartida y crear sobre ella algunas herramientas que creíamos que podían ser útiles para visibilizar y contextualizar la información estadística, como los gráficos interactivos. Porque no es lo mismo informar de que el paro sube en X personas, que poner esa subida en contexto y mostrar un gráfico con la tendencia histórica.

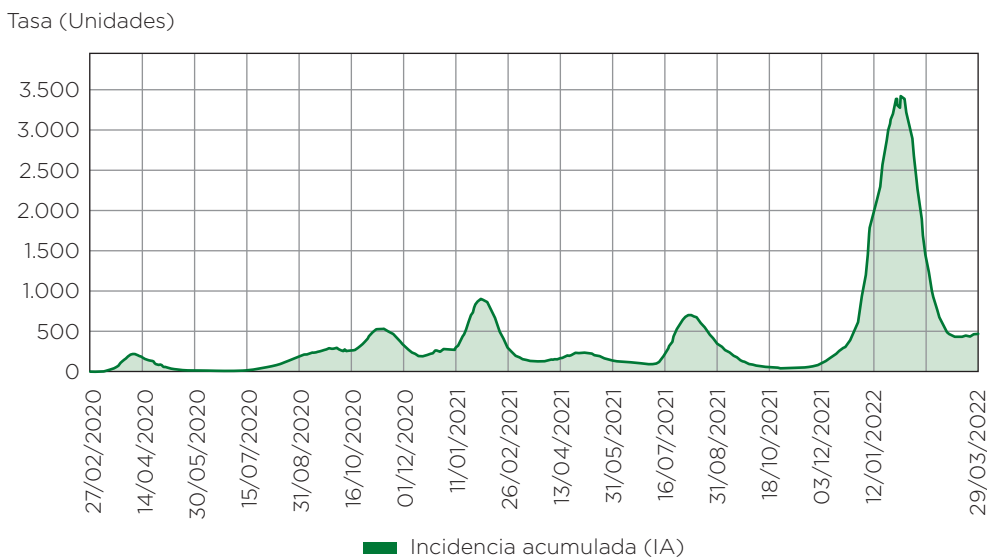
Igual que no es lo mismo decir que la incidencia sube un miércoles en 30 puntos, que ofrecer una serie de datos de libre uso para que cada uno pueda hacer sus propios análisis. En nuestro trabajo con el COVID-19 es algo que aprendimos rápido. Éramos conscientes de que los datos se veían afectados por cosas ajenas a la evolución de la pandemia. Por ejemplo, notamos que existía una estacionalidad en los datos que dependía no de las subidas y bajadas de las sucesivas olas que golpeaban al país, sino simplemente del número de pruebas que se hacían. Aprendimos que cualquier subida en la incidencia de casos iba seguida de una subida en las hospitalizaciones y en las muertes (al

menos hasta la llegada de las vacunas), tratamos de explicar, muchas veces, la diferencia entre las muertes notificadas y los fallecidos por fecha de defunción...

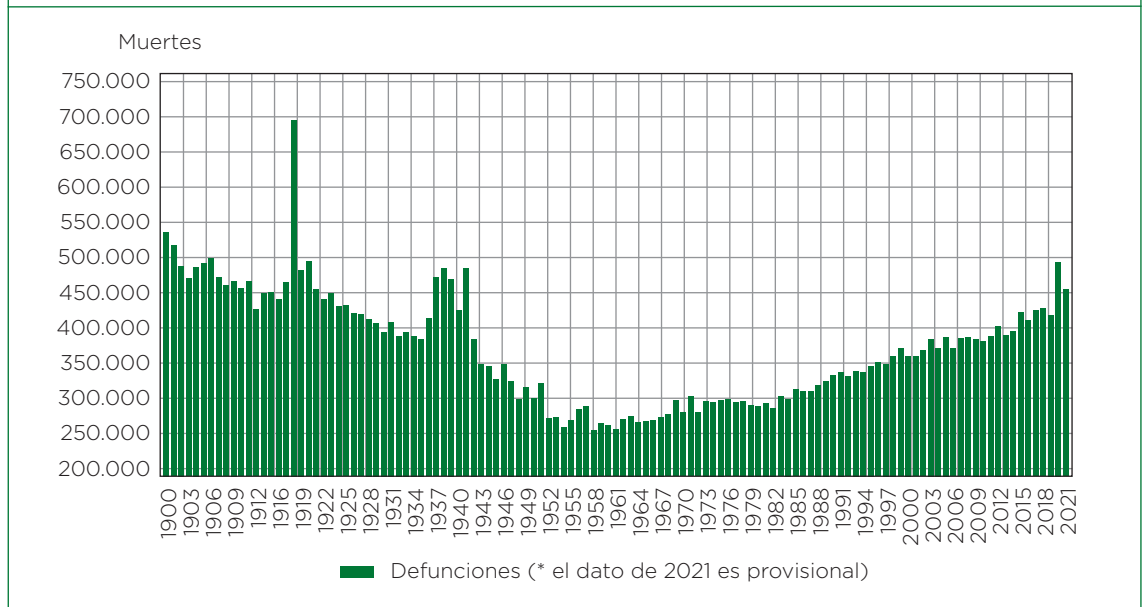
Ofrecer el dato en crudo a la gente siempre es peliagudo. “Suben los casos en 100.000”. ¿Es eso mucho? ¿Es poco? Depende del contexto, de donde se venga, de en qué punto se esté de la pandemia, del número de test... Posiblemente una de las principales virtudes del proyecto fue precisamente eso, que ofrecía contexto al lector para que él mismo decidiera. De nuevo, una premisa del periodismo de agencia: ante todo informar, sin opinar.

Casi de manera orgánica, desde las primeras series que construimos de casos y fallecidos, fuimos levantando un entramado crecientemente complejo para ofrecer más y más contexto: incidencia, número de test, casos en otros países, número de desplazamientos, muertes por cualquier causa, datos desagregados por comunidades autónomas, información local, el impacto económico, número de sanciones por saltarse el confinamiento, impacto en las residencias de ancianos... Las temáticas se multiplicaban según la pandemia iba golpeando y la administración pública refinaba su oferta de

FIGURA 1. TASA DE CASOS DE CORONAVIRUS POR CADA 100.000 HABITANTES EN LOS ÚLTIMOS 14 DÍAS EN ESPAÑA



Fuente: Ministerio de Sanidad, Servicios Sociales e Igualdad, www.epdata.es

FIGURA 2. MUERTES OCURRIDAS EN ESPAÑA DESDE 1900Fuente: INE, www.epdata.es

información. Allí dónde no llegaban los datos públicos, hacíamos solicitudes de información a Transparencia.

Teníamos modelos en los que inspirarnos. Ejemplos como el excepcional trabajo de agregación y liberación de datos realizado por 'Our World in Data' o por la Universidad Johns Hopkins, del que nos hemos beneficiado medios e

investigadores de todo el mundo. La pandemia ha servido para arrojar luz sobre las carencias en los sistemas de recopilación de datos de algunos países y también para mostrarnos el camino hacia sistemas de información pública mucho más detallados y en formato reutilizable, que se benefician de las nuevas tecnologías y del ambiente colaborativo propio de Internet para enriquecer y mejorar la capacidad de la opinión pública de fiscalizar a los gobiernos y empresas privadas.

Nada tiene que ver el primer balance que Sanidad ofreció allá por febrero de 2020 sobre el coronavirus con el número 600 por el que va ya. La información ha mejorado mucho. El Instituto Nacional de Estadística (INE) ha hecho un esfuerzo tremendo, que desde el mundo del periodismo se agradece, por ofrecer nuevas estadísticas experimentales, lanzadas en tiempos inimaginables antes de que surgiera la epidemia, con la intención de mejorar la transparencia sobre la enfermedad. Pero todavía queda camino de mejora. En EpData compartimos esa apuesta por la transparencia, por seguir aprendiendo. Y ahora que parece que la pandemia ha tomado otro curso, seguiremos adaptándonos, aprendiendo, enfocados en el objetivo de ofrecer la mejor información, en pro de la transparencia. ●

Para saber más...

Algunos datos sobre EpData:

- Fue creado en 2017 como la apuesta por el periodismo de datos de EpData.
- Contiene más de 60 millones de datos de cientos de fuentes públicas y privadas en temáticas tan diversas como empleo, población, violencia de género, salud o resultados empresariales entre otras muchas.
- La información estadística en EpData está actualizada y validada por un grupo de periodistas de la agencia de noticias Europa Press.
- Todas las estadísticas agrupadas en EpData son accesibles a través de un registro gratuito.