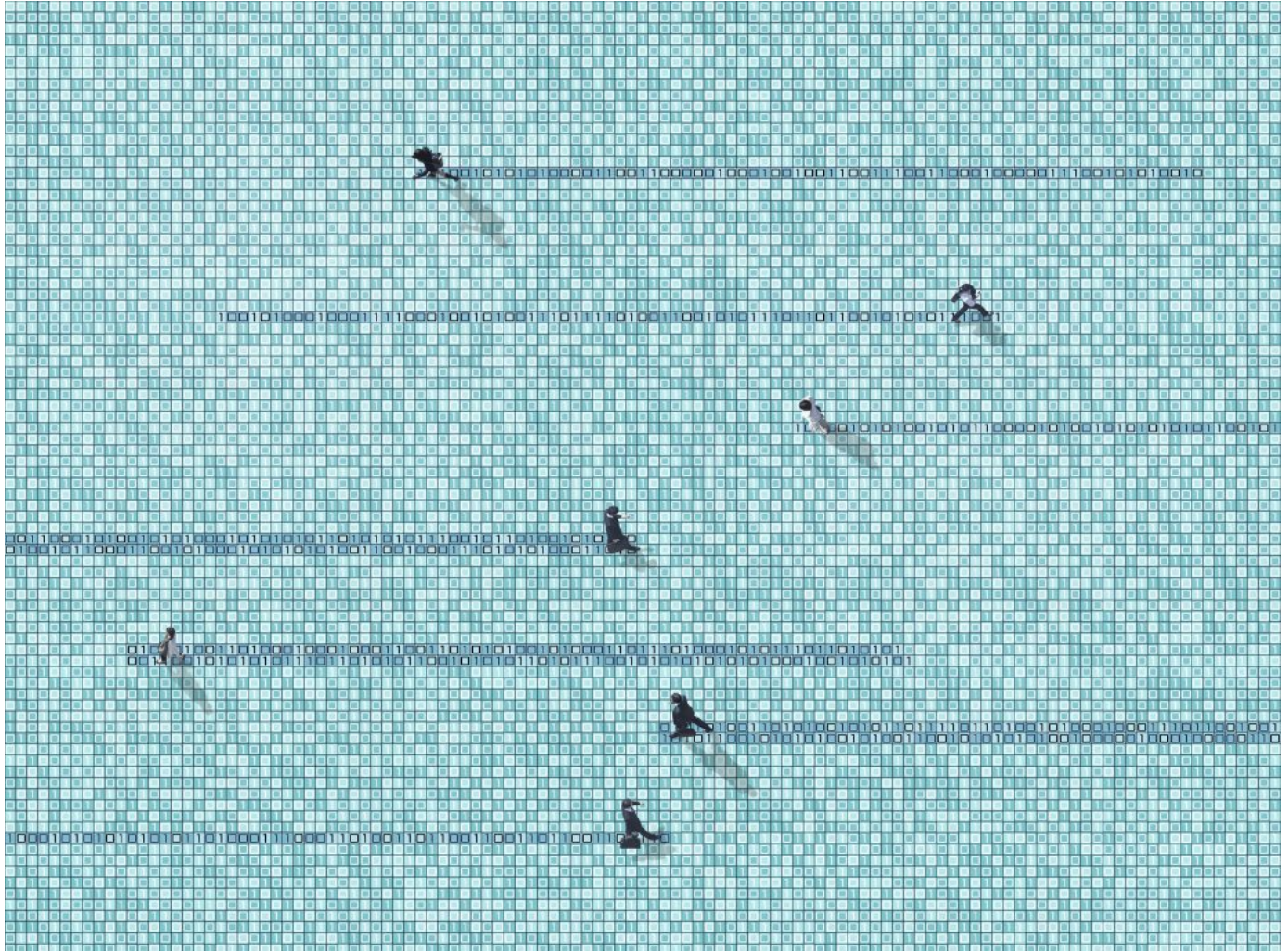


# Datos de la gente para la gente

[Álvaro Ortiz](#)



Getty Images/ Creative

***Las nuevas herramientas y las inmensas cantidades de datos en código abierto están democratizando la revolución del big data, pero ¿qué posibilidades y qué límites tiene esta transformación?***

“Cómo deben utilizar los espías, los soldados y la población la inteligencia de código abierto”. Este es el título de un reciente artículo publicado en la publicación británica *The Economist* que analiza la repercusión del gran volumen de información que recibimos sobre la guerra de Ucrania a través de “vídeos de teléfonos móviles, imágenes captadas por drones, imágenes de satélite y otras formas de inteligencia de código abierto (OSINT por sus siglas en inglés)”.

Los principales objetivos de la OSINT son descubrir, recopilar y evaluar datos que son públicos,

además de desarrollar planes y herramientas para reunir toda esa información desestructurada y procesarla. No es extraño que sea tan atractiva para gobiernos, servicios de inteligencia, departamentos oficiales de inversión, organizaciones internacionales e investigadores. Por supuesto, para esas oficinas no es nueva la idea de recopilar y analizar información, pero hay una serie de características, como el volumen, las fuentes, la capacidad de procesarla y su naturaleza de código abierto, que han empezado a transformar el terreno de juego.

El primer cambio importante es el volumen. No solo hay ya una inmensa cantidad de información disponible, sino que cada segundo se generan nuevos datos. No tiene mucho sentido calcular el tamaño en términos de *zettabytes* de Internet, puesto que el límite no para de alejarse. De hecho, ya en 2017, BloombergNEF pronosticó que el mercado mundial de almacenamiento de datos se multiplicaría por  $2^6$  (2 elevado a 6) de entonces a 2030. Así que imaginemos, ahora que estamos a punto de disponer de la web 3.0, lo que nos aguarda con los nuevos bytes de información generados por todos los dispositivos del internet de las cosas.

El segundo cambio radical que soma en el horizonte es sobre qué consideramos información. El nuevo siglo ofreció muchas posibilidades a la OSINT y el intercambio de información. La web colaborativa, la web 2.0, permitía que la gente participara y compartiera y que datos que antes no tenían relevancia de repente fueran fundamentales y empezaran a aumentar de manera exponencial. Me refiero a comentarios, conversaciones, información en redes sociales y profesionales, fotos, vídeos, audios e imágenes de satélite, además datos de empresas relacionados con las transacciones económicas y el comercio electrónico. Por si fuera poco, la mayoría de estos nuevos bytes de información tienen una referencia geográfica muy precisa.

Aunque estas nuevas fuentes generan grandes cantidades y tipos de datos, estos no están *estructurados* ni listos para su uso. Y ahí interviene el tercer cambio radical. Ha habido un rápido aumento de la capacidad de procesamiento de los ordenadores y el desarrollo de nuevos algoritmos que nos permiten convertir esos nuevos datos en información estructurada que se puede procesar y analizar. Los nuevos algoritmos que convierten texto, imágenes, audio y vídeo en números mejoran cada año, desde los más básicos hasta los algoritmos de aprendizaje profundo más complejos. Basta ver de lo que es capaz ChatGPT.

Por último, es muy importante también el hecho de que la información sea de código abierto: son datos de la gente y para la gente. Los gobiernos, las organizaciones internacionales, los organismos estadísticos y las instituciones privadas, entre otros, están trabajando cada vez mejor para que toda esa información sea pública. Además, muchas herramientas necesarias para procesarla (Python, por ejemplo) también son de código abierto, es decir, que hay una información muy abundante y unas herramientas baratas capaces de procesarla. El resultado

---

es una verdadera democratización de los datos.

Todos podemos estar de acuerdo en que el volumen y la facilidad de acceso a los datos son enormes, pero sigue habiendo interrogantes importantes: ¿cómo podemos utilizar la información con fines analíticos en ciencias sociales como la economía, la política y las relaciones internacionales y qué límites debemos ponernos? Como profesional acostumbrado a emplear técnicas de *big data* en el análisis económico, social y geopolítico, mi respuesta a la primera pregunta es fácil: las posibilidades que ofrecen las informaciones de código abierto son inmensas. No me extrañaría que la investigación en ciencias sociales empezara a adaptarse a este nuevo paradigma haciendo hincapié en la investigación empírica y experimental y trabajando en equipos más interdisciplinarios.

Cada vez hay más tendencia a utilizar estas nuevas técnicas en el campo de la economía, pese a las reticencias iniciales al uso de la inteligencia artificial (IA) y los macrodatos, hace unos años. La covid-19 mostró al mundo parte del valor de las técnicas de *big data*. La magnitud y velocidad de transmisión de la pandemia hizo que los investigadores se centraran en los datos masivos de alta frecuencia que ayudasen a entender lo que estaba ocurriendo sobre la marcha y en alta definición. Se pusieron los nuevos datos procedentes de la movilidad y las transacciones de las instituciones financieras al servicio del bien público como complemento de las estadísticas tradicionales.

Además de la capacidad de los macrodatos de mostrar lo que sucede mientras sucede, también tienen enormes posibilidades de aplicación en otros tipos de análisis. El carácter tan detallado y la alta definición de estos nuevos tipos de información mejorará la capacidad de analizar la heterogeneidad de los individuos. Estos nuevos datos son fundamentales para comprender la magnitud y el origen de las desigualdades, los efectos de la inflación en cada hogar y la repercusión de las políticas en cada casa y cada empresa. Además, esta nueva información estimulará sin duda el análisis geográfico, sectorial, urbano y del cambio climático. Sobre todo, en el caso del análisis de sostenibilidad y cambio climático, en el que disponer de datos precisos es crucial para afrontar los problemas relacionados.

Como es evidente que una solución concreta no siempre vale para todo, disponer de una información tan precisa permitirá diseñar políticas inteligentes, en función de dónde sean más necesarias y puedan tener más efecto.



La política y las relaciones internacionales también son ámbitos en cuyos análisis pueden ser muy útiles las nuevas fuentes de información y herramientas. La revolución del *big data* y la capacidad de convertir en números el texto de noticias, vídeos, documentos oficiales, imágenes y mapas pueden mejorar la calidad del análisis. Es más, estas nuevas fuentes son cruciales para estas disciplinas, sobre todo por las posibilidades de información en tiempo real. Por ejemplo, en 2011, nos enteramos de los acontecimientos de El Cairo a medida que sucedían gracias a las redes sociales y especialmente a ciertas cuentas de Twitter. En 2015 seguimos con gran precisión el tráfico de inmigrantes de Siria a Europa sin tener que hacer nada más que extraer y procesar las informaciones que aparecían en las noticias. Ahora bien, la cantidad de imágenes que están llegando procedentes de Ucrania desde que empezó la guerra no tiene precedentes.



Esta nueva oleada de algoritmos de IA nos ayuda a ver lo que está ocurriendo y al mismo tiempo a entender qué siente el mundo al respecto, el “cómo está ocurriendo”. Además de permitirnos estar al tanto en directo, la información contenida en las noticias se puede utilizar para analizar los sentimientos y quizá para medir conceptos abstractos como la incertidumbre, la felicidad y la insatisfacción a partir de lo que figura en discursos y documentos e incluso lo que dicen las personas. Entender lo que dicen los gobiernos, cómo lo dicen y cómo reacciona la población es mucho más fácil que nunca.

Dos factores fundamentales son la velocidad y la capacidad de procesamiento. Por ejemplo, hace años se habría tardado mucho tiempo en analizar los documentos que salían de las reuniones del Partido Comunista Chino, pero ahora existen técnicas de procesamiento del lenguaje natural con las que podemos acceder a esos documentos, limpiarlos y clasificarlos por temas, relevancia y los nexos entre las políticas y estrategias, todo ello con relativa rapidez. Más aún, el análisis de redes puede ayudarnos a comprender mejor las relaciones entre personas, organizaciones, políticas y países. En un mundo en el que no deja de aumentar la incertidumbre, comprender cómo interactúa un país, un sector o un líder con los demás puede ser crucial para los investigadores y los profesionales.

¿Y qué límites tienen toda esta información y estos algoritmos? En principio, es fácil imaginar que el volumen de información seguirá aumentando exponencialmente y que gran parte de ella se incorporará a plataformas de administración electrónica y otras formas de información de

código abierto u OSINT. Los límites y los interrogantes planteados se refieren más a la forma que tienen los gobiernos de regular la información y a la manera de garantizar el equilibrio entre la eficacia y los aspectos legales y éticos. De hecho, este es un problema estratégico que están abordando a distintas velocidades Estados Unidos, la UE y China.

Las posibilidades seguirán aumentando. Una mayor potencia de procesamiento, los nuevos ordenadores cuánticos y otros avances permitirán procesar fenómenos más rápidos y complejos. Los algoritmos observan y reproducen muy bien los patrones y van a hacerlo cada vez mejor en estos primeros peldaños de la escala de causalidad. Se les da muy bien ver y señalar los datos estadísticos y las correlaciones y reproducen los patrones con más eficacia que los humanos. Saben responder sobre qué pasó y describirlo. Si preguntamos a ChatGPT sobre el pasado en materia de codificación, historia o política, la respuesta nos sorprenderá.

Sin embargo, hay margen para mejorar en el aspecto de la causalidad y especialmente el de la elucubración. Una cosa es detallar lo que pasó y otra, muy distinta, explicar por qué pasó (inferencia causal), o elucubrar (sobre qué habría pasado si, por ejemplo, la Unión Soviética no hubiera caído a principios de los 90), o analizar cómo debería ser el mundo. Estas preguntas básicas son fundamentales para hacer previsiones, simular situaciones hipotéticas y diseñar buenas políticas.

Las nuevas tecnologías, las inmensas cantidades de información, las nuevas fuentes de datos y las herramientas más sofisticadas no van a desaparecer, y a partir de ahora van a transformar nuestra manera de abordar los problemas y la investigación en ciencias sociales, lo que redundará en beneficio de otras muchas disciplinas. Por supuesto, las oportunidades implican también obstáculos y para superarlos tendremos que cambiar nuestras aptitudes y nuestra forma de trabajar, desarrollar unas habilidades que nos permitan distinguir entre información y desinformación, entender los datos y su sesgo y cómo afrontarlos. Para sumarnos por completo a esta revolución y hacer realidad todas sus posibilidades, necesitamos equipos multidisciplinares y una colaboración cada vez más estrecha entre los gobiernos, las empresas y el mundo académico.

*La versión original y en inglés fue publicada con anterioridad en [Insights](#). Traducción de María Luisa Rodríguez Tapia.*

**Fecha de creación**

9 febrero, 2023