

Clivajes
Revista de Ciencias Sociales

Clivajes. Revista de Ciencias Sociales

ISSN: 2395-9495

<http://cort.as/-9CKk>

IIH-S, UV, México

Héctor Alberto Hoz Morales

MAYER-SCHÖNBERGER, V. Y CUKIER, K. (2013). *BIG DATA. LA REVOLUCIÓN DE LOS DATOS MASIVOS*
Clivajes. Revista de Ciencias Sociales. Año V, número 9, enero-junio 2018, pp. 189-194.

Instituto de Investigaciones Histórico-Sociales

Universidad Veracruzana. México

Recibido: 19-10-2017

Aceptado: 30-10-2017

Publicado: 01-01-2018

MAYER-SCHÖNBERGER, V. Y CUKIER, K. (2013). *BIG DATA. LA REVOLUCIÓN DE LOS DATOS MASIVOS*. MADRID: TURNER PUBLICACIONES.

Héctor Alberto Hoz Morales
Universidad Veracruzana
México

“Dios ha muerto”. En diversas ocasiones, la muerte de la divinidad, proclamada por Hegel y Nietzsche, se ha abierto camino hacia una nueva obra. La que nos atañe, *Big Data*, proclama, bajo la misma figura metafórica, que las certezas que sostienen al mundo están cambiando (¿han cambiado ya?) una vez más, pero esta vez “por pruebas más sólidas”.

El argumento central de la obra de Viktor Mayer-Schönberger y Kenneth Cukier es de la mayor trascendencia para el conocimiento humano, en general, y para las ciencias sociales, en particular; a saber, que la aparición y el uso de los datos masivos está por provocar (y lo hace ya, de hecho, en áreas específicas del saber humano), “una revolución que transformará cómo vivimos, trabajamos, y pensamos”, frase que aparece como subtítulo de la obra en su idioma original.

Académico el primero (Oxford University) y periodista el segundo (The Economist), los autores exponen la influencia que tienen los datos masivos en ciertas áreas de la economía y el saber humano actualmente, y lo que puede esperarse en un futuro más bien cercano.

La pareja academia-periodismo constituye, de por sí, un indicio con respecto a las transformaciones en el campo del conocimiento que podrían venir de la mano con esta última revolución tecnológica.

No existe ninguna definición rigurosa de los datos masivos, o *big data*, afirman los autores. Estos deben ser entendidos en un contexto en el que el volumen de información ha aumentado de manera exponencial, lo que ha favorecido nuevas tecnologías de procesamiento que permiten recopilar y analizar “cientos de miles de millones de puntos de datos”: para el año 2007, el 93% de los datos eran ya de carácter digital, cuando al quiebre del siglo apenas una cuarta parte de éstos se encontraba bajo esta forma. Por sí solo, el hecho motiva a la reflexión, cuando la propagación en masa de la computadora personal apenas data de hace medio siglo.

El análisis de casos concretos en que los datos masivos han transformado la manera de concebir el mundo y sus problemas ocupa buena parte del texto. Desde el uso de los datos masivos en el

baseball, hasta la predicción sobre la propagación del virus H1N1, pasando por sitios *web* que ahorran enormes cantidades de dinero a los compradores de boletos de avión o el servicio de traducción de *Google*, los datos masivos transforman ya de cierta manera nuestra realidad.

La ciencia económica, desde sus orígenes como tal, ha buscado cuáles son las fuentes de valor. La naturaleza, el comercio, el plustrabajo. Estamos quizá ante una nueva forma de creación de valor que, como lo hicieron la agricultura o la revolución industrial en su momento, permitirá una transformación no sólo cuantitativa, sino cualitativa de la sociedad: los datos masivos tienen, al menos de manera potencial, la capacidad de “revolucionarlo todo, desde las empresas y las ciencias hasta la atención médica, la administración, la educación, la economía, las humanidades y todos los demás aspectos de la sociedad”.

Entre esos otros aspectos de la sociedad, está el que nos atañe. Las ciencias sociales se verán profundamente transformadas, según la perspectiva de los autores de esta obra. En mayor o menor medida, hay cuestiones que se dan por hecho en el trabajo del científico social. Existe una limitación *a priori* en todo trabajo de investigación: no podemos aprehender la realidad y sus

complejidades en su totalidad. ¿Y si hubiera una herramienta que nos lo permitiera? Es esta la posibilidad planteada en *Big Data...*, y sus implicaciones para las ciencias sociales son, en efecto, revolucionarias.

El propio capitulado del libro no deja más que entrever la perspectiva pragmática de los autores, desarrollada a lo largo de la obra. Los primeros cuatro capítulos (titulados sucintamente “Ahora”, “Más”, “Confusión” y “Correlación”) parten de una breve descripción del estado actual de los datos masivos, y contienen los aspectos más relevantes en cuanto concierne a la ciencia social.

La revolución del *big data* trae consigo tres grandes cambios, cada uno de ellos cuidadosamente expuestos a partir del segundo capítulo. La importancia de estos tres cambios, siguiendo a los autores, es fundamental, al trastocar todos los principios de la ciencia hasta la actualidad. Acostumbrados a un mundo donde la información era escasa, el muestreo se convirtió en la herramienta privilegiada de la investigación científica, con sus inherentes limitaciones. En un mundo de datos masivos podemos, empero, trabajar con la *totalidad* de los datos de un fenómeno dado. Esto permite una mayor libertad para “dejar que los datos hablen”, que muestren diferentes

perspectivas, que den luz sobre anomalías antes no detectadas, que permitan analizar subcategorías, y, sobre todo, que puedan ser recogidos sin sesgo alguno, de manera pasiva, ya que los agentes que los proporcionan no son, en buena medida, conscientes de que lo hacen. Pasar de *algunos* datos a *todos* es el primer cambio.

De igual manera, el hecho de contar con *todos* los datos de un fenómeno, en específico, nos permite sacrificar la exactitud pretendida hasta nuestros días, a cambio de una mayor profundidad de análisis. Esto implica una transición de la exactitud hacia la probabilidad, lo que en la mayor parte de los casos es suficiente: el servicio de traductor de *Google* es, simple y sencillamente, el mejor que existe. Y lo es no en virtud de su exactitud, sino por la probabilidad de dar traducciones precisas alimentado por *todos* los datos disponibles. Aceptar la imprecisión y el desorden y abandonar así las pretensiones de exactitud es el segundo cambio.

Consecuencia de estos dos, se propone una última transformación que puede ser, en sí misma, revolucionaria, y rompe con los paradigmas más arraigados en nuestro trabajo: el abandono de una pretensión de búsqueda de causalidades, costosa y a menudo ineficiente, y su sustitución por las

correlaciones existentes entre datos. Esta transición de explicaciones causales hacia explicaciones matemáticas, probabilísticas, en las que la correlación entre variables juega un papel central, aparece cuando lo importante son los *qué*, mientras que los *porqués* pueden quedar relegados a un segundo plano. Saber *qué* causa el cáncer (o la violencia social) es más importante, en la medida en la que se puede responder a ello, a saber el *porqué*. La natural curiosidad humana por establecer explicaciones causales no podrá ser del todo desechada, pero sí dejará de tener la relevancia que mantiene cuando de resolver situaciones problemáticas se trate.

¿Se trata, entonces, del “fin de la teoría”? En el sentido tradicional, las investigaciones son guiadas por hipótesis, derivadas a su vez de modelos explicativos con distintos grados de complejidad. En el mundo del *big data*, estas tenderán a tener un rol disminuido, cuando sean precisamente los datos masivos los que orienten las investigaciones. Se puede prescindir de hipótesis, sesgadas, insuficientes y causales, cuando se dispone de *todos* los datos y de las herramientas para analizarlos, cuyo tratamiento pertinente nos podrá llevar a descubrir correlaciones suficientes entre variables

no para explicar un problema, sino mejor aún, para actuar sobre él.

El quinto capítulo de la obra, *Datificación*, parte de una premisa aparentemente simple: los mayores logros de la humanidad provienen de nuestra capacidad de medir y cuantificar el mundo. Desde Mesopotamia hasta Facebook, hemos cuantificado la realidad que nos rodea. El mundo físico primero, las palabras y la localización después, hasta la datificación de nuestras interacciones más íntimas. “Facebook datificó las relaciones, (...) Twitter (...) los sentimientos”.

La datificación del *todo* permite a su vez hablar de los datos como una nueva fuente de valor económico, objeto del sexto capítulo. Los datos sobre el mundo tienen, en la medida que son objetos de uso, un valor por sí mismos, mientras que empiezan a parecer “un nuevo factor de producción”. Se trata de bienes no-rivales. Su uso por alguien no excluye a alguien más de utilizarlos. Pueden ser explotados con propósitos múltiples, y en múltiples ocasiones, son reutilizables. Su valor se potencia al combinarse con otros conjuntos de datos, e incluso los “desechos” digitales tienen valor. La valoración global de Facebook (una compañía que no es “nada más que datos”) es de 104 mil millones de dólares, mientras que activos son de apenas 6,300 millones de dólares: el

ejemplo perfecto de una nueva forma de valorización.

Las implicaciones de los datos masivos en el sector empresarial son analizadas en el capítulo siete: se han transformado ya los modelos de negocios, han surgido empresas dedicadas a la intermediación de los datos, los criterios de decisión empresarial tienden ahora a ser informados por el análisis de datos en detrimento de expertos en el área, y se han generado nuevas ventajas competitivas.

El lado oscuro de los datos, así llamado por los autores, es abordado en el octavo capítulo, *Riesgos*, mientras que en el noveno, *Control*, se sugieren los posibles cambios que deberán ser emprendidos para así evitar y mitigar los potenciales daños que conlleva el uso de datos masivos. Los riesgos identificados por los autores van en dos sentidos: el primero, la inherente pérdida de privacidad que implica vivir en un mundo en el que cada relación y decisión del individuo es cuantificada: la vigilancia se ha vuelto “más fácil, más económica y más poderosa”. El segundo, anuncia una tétrica posibilidad (al estilo *Minority Report*), en la que la probabilidad y las correlaciones se convierten en una herramienta para prejulgar y establecer perfiles criminales. Se vuelven imperativos entonces tres modificaciones

en los paradigmas actuales de la convivencia humana: si la privacidad es cosa del pasado, será necesario garantizar que el uso de los datos masivos se haga siempre con responsabilidad, dando lugar así a un nuevo modelo de profesionales bautizados como “algoritmistas”, es decir, aquellos expertos que deberán garantizar el buen uso de las enormes cantidades de información existentes.

A partir de ahora (título del último capítulo), la sociedad de la información “cumple con la promesa implícita en su nombre”. Toda la información disponible es susceptible de ser organizada y analizada de una y mil formas, que permitan crear valor económico o valor intelectual. Los datos masivos son, en este sentido, una herramienta más, cuyo buen uso estará en función de nuestra capacidad de adaptar una “mentalidad *big data*”: capacidades técnicas y creativas de la mano de un sentido de responsabilidad que garantice la privacidad e individualidad humanas.

De más está decir que la obra aquí reseñada no es, ni pretende ser, un tratado de metodología. No obstante, ofrece una seria oportunidad para emprender una reflexión formal y sistemática acerca de los retos y desafíos que supone la realidad de los datos masivos.

Será necesario, en este sentido, un fino tratamiento epistemológico para discernir, en un primer momento, si los cambios señalados en los primeros capítulos tienen una significación tal para las ciencias sociales como lo han tenido para otras áreas del conocimiento, y en segundo, si es posible, deseable o necesario llevar a cabo una defensa de la causalidad ante las explicaciones probabilísticas, o si, efectivamente, “Dios ha muerto”.