

## PAPEL CRÍTICO 101

Javier Callejo\*

Universidad Nacional de Educación a Distancia (UNED)

### Big Data. Topología e Investigación Social

Autores: Fernando Conde Gutiérrez del Álamo

Páginas: 333

Editorial: Editorial UNED, 2023

Ciudad: Madrid

Como viene ocurriendo con otros fenómenos e innovaciones —globalización, crisis del Estado, polarización social, criptomonedas, inteligencia artificial, etc.— el análisis sociológico tiene una relación ambivalente con los denominados big data. Por un lado, gran amenaza, porque aquí parece que todo es grande, que deja la metodología de las ciencias sociales en situación zombi, como plantean Savage y Burrows. Al menos, amenaza para una forma de hacer sociología, que es más o menos la que se viene realizando desde el final de la Segunda Guerra Mundial. Por otro, gran esperanza para la recuperación de la disciplina como fuente de los grandes y profundos diagnósticos sobre los procesos sociales. Repito que, cuando se habla de los big data, todo toma otras dimensiones. Pues bien, el libro escrito por Fernando Conde es un reflexivo e interesante intento de arrebatar los big data a la exclusiva gestión de ingenieros, incluidos los ingenieros sociales. Así, puede leerse el texto, en especial su última parte, como un enfrentamiento entre Conde y Comte. Pero hasta llegar allí quedan párrafos de interpretación de un texto importante, que, sobre todo, aporta claridad sobre el sentido sociológico de los big data. Tal es su objetivo. Por si no queda claro, una advertencia: no se trata de un libro de instrucciones sobre las operaciones de gestión de grandes bases de datos para fines de investigación sociológica. Es una interpretación sociológica, que incluye la investigación social empírica, de los big data.

La obra se enfrenta a la pregunta sobre qué será de las ciencias sociales en la época de los big data. Se responderá articulando una rebaja de las expectativas puestas en los mismos para explicar, ordenar y prescribir el mundo social; con una recuperación de la fuerza de la visión o, para decirlo con mayor apego al texto, de la teoría y las formas de hacer sociológicas.

\* **Correspondencia a / Correspondence to:** Javier Callejo. UNED, Facultad de Ciencias Políticas y Sociología, Departamento de Sociología I (Teoría, Metodología y Cambio Social). C/del Obispo de Trejo, 2 (28040 Madrid) – [mcallejo@poli.uned.es](mailto:mcallejo@poli.uned.es) – <http://orcid.org/0000-0002-0856-5642>.



Al menos de una forma de hacer teoría y practicar la sociología. De esta manera, nos acercamos a otra pregunta de igual interés que la que dirige el libro, pero que parece impertinente dados los tiempos que corren, como es: qué será de los big data en la época de las ciencias sociales. Sin embargo, creo que el libro, en su esfuerzo de domesticar la comprensión de los big data, tiene también mucho de respuesta a esta pregunta; aunque no se la plantee de forma directa.

De la mano de Latour y Venturini, autores con los que se mantendrá en continuo diálogo a lo largo de todo el texto, Conde propone los datos masivos —por cantidad, que no por desorden— como el gran disolvente de dualidades, fijándolos en el nivel de lo analítico. Para ello, los expulsa del nivel ontológico. Incluso en ese nivel analítico tendrán que buscar aliados en métodos y técnicas de la investigación social que ya estaban ahí, antes de que llegasen. Una alianza o articulación que tendrá un espacio. Es decir, algo en el que cada cosa ocupa un lugar más o menos estable, con formas inalterables y con fronteras. Es el espacio de la disciplina de los espacios, el espacio topológico. Pero esto configura la tercera parte del libro. Empecemos por el principio.

La fiesta del texto se introduce en la respuesta a esa pregunta existencial sobre el devenir de la sociología en tiempos de datos masivos con fuegos artificiales. Con la polémica sobre la digitalización en las ciencias sociales, una polémica muy francesa y, por tanto, tendente al infinito. Tal vez era necesario pasar por ella, pues buena parte de las obras referidas parecen estar inscritas en la misma. La segunda cuestión que se aborda, vinculada a la primera, es si los datos masivos son una revolución o no en las ciencias sociales. Demasiado vinculada a la primera como para tener entidad propia. Que los datos masivos abren nuevas líneas de investigación social aplicada parece claro. Que constituya una revolución o una especie de nuevo paradigma es más discutible. Los paradigmas suelen venir cargados de teoría y el paradigma de los *bigdatáfilos* o simplemente *datáfilos* tiene como teoría prescindir de la teoría. Pero está bien plantearlo, dadas las excesivas pretensiones de este movimiento. Más allá de si constituye una revolución u otra cosa, lo cierto es que la trazabilidad extendida que supone la digitalización es un paso irreversible con consecuencias para la metodología de las ciencias sociales.

La trazabilidad masiva da un vuelco a la manera de entender la metodología en las ciencias sociales. Véase, como se señala en el libro, el caso de la representatividad. De la representatividad que podemos denominar tradicional, que reposaba sobre los distintos tipos de inferencia, distributivo-estadística o estructural, tomando los términos de Jesús Ibáñez, a una representatividad que puede considerarse implícita. ¿Tiene sentido seguir hablando de representatividad en tiempos de trazabilidad, más allá de hasta qué punto el significativo registrado representa al individuo o comportamiento?

En la primera parte del libro se enfocan, con desarrollo desigual, las V que caracterizan los registros masivos: variedad, volumen y velocidad. Ya en esta primera parte deambula el texto en proyecciones de lo que la nueva digitalización puede suponer para la sociedad. Algunas, como las de la huella digital con el internet de las cosas (IoT) más fundamentadas en las apreciaciones de jóvenes autores (Schmitt), que en observaciones empíricas. Fuera de las V, pero también señalado con acierto en la segunda parte del libro, se apunta que bajo el dominio de la lógica de la trazabilidad digital todo es susceptible de convertirse en información. Hasta el ruido, el error, el usuario que desoye las recomendaciones (p. 90), se convierten en información. Hemos pasado de la imposibilidad de la no-comunicación a la concepción de la

no-información como un gasto, desvío o pérdida del análisis. La no-información es una información no observada. Mejor dicho: no trazada.

Muy interesante y pertinente es la tipología de big data que se establece: sociodemográficos, conductuales y ambientales. Se ofrecen otras clasificaciones, pero ésta es lo suficientemente clara. Menos convincente es la clasificación, ya en el capítulo 6, de las huellas digitales: caracterizadas, relacionales y aisladas, definidas estas últimas porque no requieren ir asociadas a ninguna otra característica o atributo (p. 86). Una clasificación en la que parece latir la evolución de la web —huellas web1, huellas web2 y huellas web3— pero eludiendo el significado que adquiere la web3 como la web del valor, donde las huellas —algunas de sus huellas— son directamente significante de valor.

Considero que la mayor aportación de la obra se encuentra en esta segunda parte. Cuando se enfrenta más directamente a las relaciones entre registro, dato, huella digital, registro digital, sus colaterales (producción de registros o producción de datos), excedente de información, o mismamente concepto, aun cuando esto de manera indirecta, pues parece que el término operacionalización adquiere otro significado. Tienen especial valor las indicaciones de la relación entre los elementos de este nuevo sistema conceptual, como la aparición del dato como el resultado de una operación de cálculo sobre los registros, una vez codificados (p. 62) o la descripción del paso del registro digital al dato. También es donde se encuentra la propuesta más original, aun cuando sólo sea por su sistematicidad. Las huellas son presentadas como fruto de un momento de registro; para que haya huella, tiene que haber registro, pues no parecen preceder al registro, como, por ejemplo, la huella del oso del bosque precede al registro que pueda hacer el cazador. Las huellas digitales son consecuencia del registro. Sin embargo, tal concepción parece matizada en otras partes del texto, pues el registro puede interpretarse como una operación precedida de la existencia de huellas. Así ocurre cuando se dice: «*desde la inscripción y registro de las huellas hasta la presentación visual de los datos como resultado de los análisis realizados*» (p. 104). La originalidad se encuentra en la apuesta que sitúa al registro configurando la huella y es lo que en buena medida recoge de las aportaciones de Boullier: las huellas digitales carecen de todo referente semántico (p. 109). Confundir huella física y huella digital nos llevaría al barro en el que luce primorosamente la primera. Pero estamos hablando de lo digital y no de tierra empapada.

Si el entramado conceptual constituye el alma de esta obra, el cuerpo está en la justa relevancia que pone en la visualización. Es el tema de la tercera parte.

La interpretación big data parte de gráficos. Su territorio y, podría decirse, materialidad están en mapas y grafos. Lo aparentemente evanescente, como son inmensas nubes de datos, que parecen escaparse, como el humo, se condensa en la relativa rigidez de las pocas dimensiones que pueden representarse en una hoja o, a lo sumo, una pantalla, incorporando así una tercera dimensión. El territorio de los big data está en su capacidad para reducir una hipotética infinitud de datos, relaciones y dimensiones al territorio de un mapa o grafo. Ahora bien, como puede llegar a deducirse de las reflexiones más brillantes del texto: es el mapa imaginario de los big data el que, como si fuese un aplicado seguidor de Borges, genera la sensación de estar ante un territorio. Como en los mundos imaginarios de Borges que nos hablan de enciclopedias con un realismo que roza lo tangible, los mapas de los big data adquieren valor desde su coherencia (p. 156). Tan coherentes que los hacemos reales, como si la realidad fuese eso, una fuerza incoherente domesticada en coherencia. El caos y la incoherencia de lo social y las fuerzas sociales son domesticadas y presentadas en forma de sociedad con

estables posiciones en los mapas y grafos de los big data. El problema que se abre y se remitirá hacia el último bloque del libro es a dónde nos llevan estos mapas. Mapas que nos ponen el territorio en sociedades desterritorializadas, siguiendo aquí la estela teórica abierta hace más de cincuenta años por los antiedípicos Deleuze y Guattari. Autores que tendrán un papel central en el cierre del libro, en el epílogo. Eso sí, ya descargados de psicoanálisis y las potenciales demandas de descentralización más o menos esquizofrénicas.

Es tal la importancia que adquiere la representación gráfica que se deduce un cambio en la forma de conocimiento —el nuevo, de carácter abductivo— y, en definitiva, de pensamiento. Ya decíamos que en esto del big data todo es grande y con mayúsculas. Un cambio que, sobre todo para quienes hemos crecido en un mundo más dominado por la gramática y la sintaxis argumental, que por el sintético holismo de las pantallas, requiere de una piadosa alfabetización en la lectura topográfica. Al menos, para adquirir las mínimas bases que permitan la interpretación de los resultados de los big data. Es así que el big data requiere una guía comprensible para nuestra conversión. Con una guía constituida por casos formados por investigaciones concretas, entramos en el núcleo propositivo de esta tercera parte del libro. Así, empieza con el momento de la investigación social en el mundo que vivimos, sigue con los big data y se sale con la topología. Un recorrido secuencialmente expositivo, analítico y propositivo.

Antes de entrar en la propuesta, fijemos en qué consiste esta nueva forma de pensamiento, pues es el marco en el que ha de encajar tal propuesta. De manera muy sintética, haciendo poca justicia al desarrollo expuesto en el libro: a) la primacía de la percepción sobre el razonamiento analítico; b) la inclinación hacia la lectura holística; c) la posibilidad de lecturas parciales, como esas que, a modo de narraciones de Umberto Eco, nos llevan a buscar la clave de crípticos mensajes en, por ejemplo, las esquinas de los tapices. Potenciales lecturas que, como ya nos advierte el autor, estaban ya presentes en algunos análisis multivariantes del, ahora, pensamiento estadístico tradicional, como es el escalamiento multidimensional o el análisis factorial de correspondencias múltiples. Por ello, se muestra (capítulo 9), el paso de la visualización en el análisis factorial de correspondencias múltiples a la visualización de redes complejas. Hay formas de ver compartidas; pero, siendo esto lo importante, lógicas distintas. Tal vez por ello y por esa falta de alfabetización en la interpretación de resultados que requieren otra forma de pensamiento, la relativa transparencia en la interpretación de las salidas de un escalamiento multidimensional o un análisis factorial de correspondencias múltiples se nos transforma en opacidad, cuando se compara con las salidas de los big data. Y es que, como se nos advierte en el libro, la transparencia interpretativa en las salidas del análisis big data es menor. Por ello, tiene especial relevancia el apartado titulado: «Algunos errores habituales en la lectura e interpretación de los mapas de las redes complejas» (pp. 219 y ss.). Una escueta pero significativa llamada de atención. Pues esto nos lleva a la reflexión sobre el significado del concepto de error en la interpretación-visualización de los resultados big data. La metodología surge de la lucha contra el demonio del error.

La propuesta cabalga sobre la crítica a otras propuestas, como las de Latour y Venturini; el seguimiento de la estela de otras, como el análisis de controversias y, sobre todo, la topología: espacio, pues qué va a ser la topología sino espacio, en el que se encuentran las perspectivas (tradicionales) de la investigación social —cualitativa y cuantitativa— con los big data. Extendiendo las metáforas de orden espacial, la propuesta se establece sobre un cruce de vías: análisis de controversias vía análisis de redes vía métodos cualitativos y cuan-

titativos (p. 228). Pero estas cabalgaduras son caballos de Troya a través de las cuales Conde nos hace regresar a textos suyos de 1987, donde se articulaban las perspectivas cualitativa y cuantitativa sobre la creación de: «*un cierto espacio topológico de encuentro*» (p. 238). La topología actúa como velo que permite la acumulación de analogías, del «como si...»: como si fueran idénticos cualitativo y cuantitativo. Se apunta que en las redes sociales se configura, al fin y al cabo, una especie de sistemas de discursos —más bien significantes claves o, siguiendo términos aportados por el texto, atractores semánticos— con sus ejes, de manera que su análisis puede dar resultados muy semejantes a lo que hacía el análisis producido por métodos cualitativos de investigación social (ejemplos 1 y 2 del capítulo 11).

Un texto que tiene por objetivo captar el sentido de los big data, tal vez con la sana intención de darle la vuelta, ha de hacer hueco a la retórica propositiva. Pero la retórica propositiva es un tanto peligrosa, pues, de tan escurridiza, nos hace deslizarnos a afirmaciones a veces demasiado rotundas. Como aquella que habla de: «*plena coincidencia con los resultados de la investigación cualitativa a este respecto*» (p. 170). Al hacer fuerza en lo que se escurre, ya se sabe lo que probablemente pase.

La cuarta parte se configura a modo de conclusión. Nos habla críticamente de las predicciones que prometen los *bigdatáfilos* como Pentland. Hasta los párrafos de esta cuarta parte y capítulo final, antes del epílogo, se había pasado de puntillas por el gran arma esgrimida por el *bigdatismo* contra todo método anterior: la predicción. La capacidad de predicción es el escudo de su bandera, de la reivindicación de su validez pragmática. Predicciones que son esfuerzos de persuasión, de prescripciones. En toda colonización del futuro, hay una imposición sobre el presente y, por lo tanto, de ese futuro. Pasando de poner el acento en las conexiones (redes complejas) a hacerlo en las correlaciones, Conde rechaza el rechazo de esos *bigdatáfilos* al análisis de la causalidad. Ahora bien, la negación de la negación tiende a empujarnos por caminos un tanto pringosos. Justificada la crítica a la no-teoría defendida por la *bigdatafilia*, pues una no-teoría es una teoría. Pero una cosa es la pertinente crítica al exclusivo protagonismo de la correlación, elevándola muy por encima de su estatus de sólido indicio de relaciones, y otra caer en los brazos de la causalidad.

Más que una conclusión, este capítulo final se nos aparece como un destino del texto. Lo deja claro el subtítulo: «las concepciones de fondo del abordaje dominante del big data». El destino está en el fondo. Un fondo oscuro. Casi apocalíptico. A partir de aquí, los big data no informan o no sólo nos informan, sino que nos conforman, pues llevan a los sujetos a la pérdida de consciencia (p. 277). Es el apartado de tono más crítico, que se completa en unas últimas páginas, presentadas como epílogo, dedicadas al lugar de los big data en las sociedades de control. No puede dejar de resultar paradójico que, desde una disciplina que ha hecho hincapié en poner entre paréntesis la voluntad humana individualizada (desde Marx a Bourdieu, pasando por Durkheim), se reivindique el —tan cristiano y/o existencialista— libre albedrío (p. 278). Es la salida que se da al cerrado orden de las máquinas predictivas.

En las predicciones sustantivas y más acá —o allá— de esa sociedad de control, el libro se introduce en el campo del consumo, del que el autor es referencia en la sociología española. Se presenta una sociedad de consumo *offline* agitada, controlada y motivada desde las observaciones y acciones *online*. En la que los datos producidos en internet se vuelcan para generar comportamientos *offline*. Lo *online*, para resumir, seguiría actuando como un medio de comunicación que, ahora, nos dice los intereses de la audiencia a partir del comportamiento como tal audiencia en la red. En gran medida, esto sigue siendo así. Sin embargo, especial-

mente para los sectores de población más jóvenes —los etiquetados como nativos digitales, aun cuando creo que se queda corta la etiqueta— buena parte de su consumo y, en cierta medida, su sociedad de consumo es ya predominantemente *online*: toman un cuerpo-avataar, se visten y adquieren complementos-instrumentos y toda una variedad de *skins* para sus experiencias de inmersión en la red. Esto, desde la perspectiva crítica asumida por el texto, creo que ha de tenerse en cuenta en la medida que apunta a una especie de virtual doble cierre en el mundo virtual. Observaciones en el mundo virtual para comportamientos en el mundo virtual.

*Big data, topología e investigación social* es un texto importante, que nos sitúa como sociólogos en el momento en el que los datos masivos parecen haberse apropiado del campo metodológico y, lo que profesionalmente es más importante, de la investigación social aplicada. Su cuerpo conceptual es especialmente recomendable para quienes se insertan en la sociología digital. Con la advertencia de que cada vez queda menos hueco para otro tipo de sociología.

No cabe duda de que la trazabilidad digital y los datos masivos derivados de ella están aquí retando técnica, epistemológica y éticamente a las ciencias sociales. Más allá de pretendidos profundos debates —¿son una revolución?— la virtud del texto es que nos lleva a cuestiones sociológica y socialmente muy relevantes. En primer lugar un asunto que no puede remitirse a la carpeta de problemas meramente prácticos, de la práctica profesional. Es el de la propia accesibilidad a los datos: ¿quién dispone de datos masivos? ¿qué accesibilidad hay a los mismos por parte de los investigadores externos a su diseño y producción? Producir datos es caro y la lógica de las empresas que producen los datos no es ponerlos en común, como se subraya. La otra cuestión es la de la privacidad. Ahora bien, se pone más el acento en el problema de la privacidad cuando el agente son redes corporativas, que cuando el agente es el Estado, salvo en las últimas páginas del libro. Un trato desigual, cuando especialmente el que tiende al control absoluto de los movimientos de los ciudadanos es el Estado, especialmente bajo argumentos como la gestión de la fiscalidad, la lucha contra el blanqueo de dinero o contra el terrorismo. De hecho, el actual debate bajo las intenciones de distintos países de generar sus CBDC (monedas digitales de los bancos centrales), como ocurre con el euro digital, tiene su centro aquí: la capacidad del Estado de acabar con la privacidad de todo tipo de transacción comercial. El Estado dispondría de la gran base de datos de todas las transacciones económicas realizadas por todos sus ciudadanos. Sin embargo, son las grandes empresas las que se sientan en el banquillo de los acusados de acabar con la privacidad y como mayores responsables de la configuración de la sociedad de control.

La obligación de síntesis de toda reseña no puede dejarse de subrayar que, a pesar de ser un texto profundo, se acompaña con ejemplos de investigaciones sociales empíricas traídos de la trayectoria profesional del autor. Algo que ayuda a su comprensión. Seguramente más que esta reseña. Pero permitan la justificación de que hubiera sido excesivo traer aquí el pormenorizado análisis y comentario de los casos presentados. Por otro lado, merece la pena que cada uno lo haga por sí mismo y, así, se sumerja en un libro lleno de valiosas reflexiones.