

Reflexiones epistemológicas sobre Big Data

Ignacio Carlos Maestro Cano

Universidad de Valencia/UNED

1. Exordio en torno a la naturaleza de los datos

Una cuestión previa y nada trivial que es preciso abordar al analizar lo que en los últimos años viene denominándose *big data* es el de la propia naturaleza de los datos, dado que la cuestión se centra en cuál sea el valor intrínseco de éstos. Esta importante tarea es la que este primer apartado pretenderá abordar.

Un aspecto que tradicionalmente se considera al evaluar la esencia de los datos, es el de la conexión que exista entre lo sensible y lo racional o, dicho de otro modo, entre cómo se percibe el mundo y cómo se piensa éste. De este modo, el resultado de la participación de ambas realidades del ser humano como sujeto cognoscente determinará en gran medida el tratamiento que deba admitirse para los datos. Así, se dice que el mundo le es inalienable a la conciencia, que hombre y mundo no existen como entes aislados, siendo por ello que ya Hegel (1807, p. 385) propuso que “la individualidad es lo que su mundo sea en cuanto que es el suyo”. Así, el equilibrio entre mundo *percibido* y mundo *entendido* ha constituido siempre una de las cuestiones centrales de la filosofía y la epistemología. Desde el cartesiano “pienso luego existo” (Descartes 1637, p. 68) —entendible como un “yo *me doy* con el pensamiento”, de modo que “si no interpreto, no soy”— hasta la suspensión del juicio que propusiera la fenomenología con su vuelta “a «las cosas mismas»“ (Husserl 1901, p. 7) o el existencialista “yo soy yo y mi circunstancia” (Ortega 1914, p. 30) sin olvidar, como clave de esta evolución, la figura de Kant. En este sentido, cabe destacar que, para Kant, la experiencia viene a constituir un resultado del entendimiento por cuanto “aunque todo nuestro conocimiento se eleva [*anhebt*] con la experiencia, no por eso procede todo él de la experiencia” (Kant 1781, p. 45), siendo que la “experiencia es una percepción comprendida” (Kant 1774, p. 664). Yendo más lejos, Heidegger llegaría a sostener que “para que haya en general algo conocido en un conocimiento, es decir, algo dado en general, me tiene que estar dado y ser objeto «para mí» en algún sentido” (Heidegger 1926, p. 256). Así, esta centralidad del “para mí” podría suponer que “lo percibido se *determina* en el hablar sobre ello” (Heidegger 1926, p. 15)

puesto que “el ser se determina mediante el pensamiento” (Heidegger 1926, p. 149). De este modo, la relación entre el individuo y su percepción del mundo —los datos— podría plantearse así: “El yo (...) es la condición de posibilidad de captar un ente (...) una de las cosas que enlaza es el yo, y significa: yo pienso. Yo soy aquello «para» lo que algo puede estar dado. (...) El yo es el «ser para qué» para lo dable y por tanto determinable” (Heidegger 1926, p. 261). Desde esta perspectiva, quizás discutible pero repleta de sentido, sería el ser humano quien vendría a establecer qué es lo “dable” puesto que “el dato como tal [sólo] es posible en un «para», que está constituido mediante una síntesis original que se expresa en el yo pienso” (Heidegger 1926, p. 263). Al margen de las naturales objeciones que puedan hacerse a este planteamiento, lo cierto es que es el ser humano quien hace de un mero *valor* observado un *dato*, otorgándole así cierta relevancia y con ello, por cierto, alimentando ciertas esperanzas cognitivas. Una cosa sería lo “dado”, eso que no es sino aquello que “está ahí” (un meramente-estar-ahí), y otra muy distinta aquello que, de entre lo dado, seleccionemos o convirtamos en “dato”.

La cuestión se centra en definitiva en destacar la participación del sujeto cognoscente en la determinación y valoración de lo percibido. Algo que, entrando ya en el ámbito de los datos masivos, ha sido señalado por Bowker al afirmar, en una clara referencia a la obra de Lévi-Strauss, que “los datos crudos [*raw data*] son a la vez un oxímoron y una mala idea; al contrario, los datos deberían ser cocidos con cuidado” (Bowker 2005, p. 184). En efecto, desde el momento en que el ser humano escoge la variable (establece las componentes principales de la información) y recoge los datos correspondientes, toda información queda conformada por él. Pese a ello, el apostolado *big data* tiende a considerar que, mientras los datos constituyen una entidad libre de subjetividad, el entendimiento humano siempre está expuesto a ella. Se tiende a pensar que resulta más improbable que nos “engañen” los datos a que lo haga el entendimiento humano, olvidando que los datos, pese a su nombre, distan mucho de ser algo dado sin más, presentado ante nuestros ojos de forma aséptica y neutral por un ente externo de objetividad infinita. Tal y como ha expuesto Crawford (2013), “los datos y conjuntos de datos no son [*per se*, se entiende] objetivos; son creaciones del diseño humano. Nosotros le damos a los números su voz, extraemos interferencias de ellos y delimitamos su significado a través de nuestras interpretaciones”. En este sentido, podría concluirse, ya como un primer principio a tener en cuenta, que la pretensión de una ciencia centrada exclusivamente en los datos constituye un absurdo.

Pese a lo expuesto, se asegura que no hay que obsesionarse con las causalidades sino que debemos limitarnos a “dejar hablar a los datos por sí mismos” (Mayer-Schönberger y Cukier 2013, p. 14) puesto que, “si se dispone de suficientes datos, los datos hablan por sí mismos” (Anderson 2008). Nada más lejos de la realidad. El *logos* es patrimonio del ser humano y es él el único que habla —que piensa—. Por desgracia, los datos, si hablan, lo hacen con relativa frecuencia para engañarnos y es por eso por lo que suele hablarse de la “destilación” o el “refinado” de los datos. Los datos sin el entendimiento humano —ése específico “enlazar” que caracteriza al ser humano— no aportan comprensión o, dicho a la inversa, sin entendimiento, los datos no “mientan” nada y, por tanto, no *significan* nada.

Esta oportuna reflexión epistemológica es la que podría situarse en los fundamentos de aquello que de manera muy sintética se ha expresado en “*Raw Data*” is an Oxymoron (Gitelman 2013). Así, ningún dato es un dato “puro” (*raw*) en el sentido de “neutral” o “inocente” *per se*. Puede que la naturaleza sea “pura naturaleza” (*raw nature*), pero la forma en la que la observamos —incluso en un contexto científico— difícilmente queda exenta de algún propósito y, en ese sentido, nunca podrá ser para nosotros “pura” naturaleza. Puede que una masa boscosa sea *per se* “mera” naturaleza, pero en cuanto deviene objeto de estudio para el hombre, pasa a ser una zona “a proteger”, “a aprovechar económicamente” o algo por el estilo. No se trata de ninguna comprensión materialista del hombre y su historicidad, es la simple constatación de que, cuando el hombre hace ciencia, ésta tiene un propósito. Tampoco se trata de desdeñar la ciencia básica —en absoluto— sino de indicar que no tiene cabida en este contexto el concepto de una ciencia meramente “recreativa” —lo sea consciente o inconscientemente—.

2. Encuadre epistemológico de la tecnología Big Data

Una vez revisada la importante cuestión en torno a la naturaleza de los datos, es posible analizar la problemática relativa al adecuado empleo de éstos: su vertiente metodológica o epistemológica. El empleo masivo de datos (*data-intensive research*) ha favorecido la aparición de una serie de planteamientos que debieran alertarnos de un posible exceso de confianza en la ciencia o, si se prefiere, de autoconfianza. Una mala comprensión del fenómeno *Big Data* podría agravar los problemas relativos a lo que se conoce como falta de calibración o exceso de confianza en la toma de decisiones, un fenómeno psicológico que ha

sido ampliamente estudiado. Un experimento muy difundido a este respecto es el de Alpert y Raiffa (1982). Resulta oportuno por ello recordar una serie de cuestiones epistemológicas con el fin de evitar la paradoja de que el acceso a una mayor cantidad de datos lleve a la comisión de errores más graves.

En realidad, no se trata de un problema propio de los datos masivos ni, por tanto, nuevo. Es una cuestión que ya abordara Poincaré en su obra *Ciencia e hipótesis*, en la que alertaba de la necesidad de llevar a cabo una revisión del papel de la observación empírica. Para Poincaré se trataba de poner en tela de juicio los planteamientos “que nadie puede cuestionar” y según los cuales “la experiencia es la única fuente de verdad. Por sí sola puede enseñarnos algo nuevo; por sí sola puede darnos certeza” (Poincaré 1902, p. 167). Por el contrario, Poincaré alega que “no es suficiente con observar; debemos usar nuestras observaciones, y para ello hace falta generalizar” (Poincaré 1902, p. 167), por lo que plantea la cuestión decisiva aquí:

“¿No podemos estar contentos con la experiencia por sí misma? No, esto es imposible, sería malinterpretar el verdadero carácter de la ciencia. El hombre de ciencia [*savant*] debe clasificar. La ciencia se construye con hechos, igual que una casa se construye con piedras; pero una acumulación de hechos no es una ciencia del mismo modo que un montón de piedras no es una casa” (Poincaré 1902, p. 168).

Es un trasfondo epistemológico cuya importancia en el ámbito de los datos masivos ya ha sido recogida por diversos autores, como es el caso de Berry (2011), para quien *Big Data* supone de hecho la “desestabilización de cantidades de conocimiento e información que carecen de la fuerza reguladora de la filosofía” (p. 8).

En este punto, la actitud adoptada frente al empleo intensivo de datos nos sitúa frente al dilema de posicionamientos epistemológicos como el positivismo lógico o el racionalismo crítico; ante eso que en filosofía de la ciencia se conoce como el problema de demarcación y que se ocupa, nada menos, que de establecer el límite entre lo que deba considerarse ciencia y aquello que no constituyan sino, en palabras de Popper (1934, p. 34), meros “sistemas «metafísicos»“. En este sentido, al establecer la llamada “falsación” como criterio de demarcación, Popper vino a relativizar la importancia de la observación empírica (de los datos) en la confirmación de una teoría científica ya que, en su opinión, tal observación jamás

permitirá demostrar o validar una teoría, sino que únicamente servirá para refutarla. Es más, para Kuhn (1962), y dentro eso que él denominara la ciencia “normal” (pp. 70-71) —aquella firmemente asentada sobre el paradigma existente—, cualquier cambio de paradigma constituye una decisión que “sólo puede tomarse con fe” (p. 278), esto es, un “compromiso” (p. 198) o “experiencia de conversión que no se puede forzar” (p. 269). Por ello Kuhn considera que el falsacionismo de Popper constituye una simplificación excesiva pues “las experiencias anómalas [aquellas que introducen una crisis en el paradigma científico aceptado] no pueden [meramente] identificarse con las falsadoras” y “la competencia entre paradigmas no es el tipo de batalla que se pueda resolver mediante pruebas” (p. 262).

En lo que Popper y Kuhn parecen coincidir, y que constituye un tema central en la cuestión del potencial de *Big Data*, es en que nunca se puede ser totalmente objetivo (racionalismo crítico) resultando que “no existe nada que pueda llamarse inducción” (Popper 1934, p. 39), por lo que la cuestión ha de centrarse necesariamente sobre los métodos para poner a prueba nuestro conocimiento. “Nuestra ciencia no es conocimiento (*epistēmē*): nunca puede pretender que ha alcanzado la verdad (...) No sabemos: sólo podemos adivinar” (Popper 1934, p. 259). De este modo, y puesto que “ni la observación ni la razón son autoridades” (Popper 1963, p. 235), el único método fiable para Popper es el que denomina de “conjeturas y refutaciones” reconociendo, eso sí, que “aunque creo que en la historia de la ciencia es siempre la teoría y no el experimento, la idea y no la observación, lo que abre paso a nuevos conocimientos, creo también que es siempre el experimento lo que nos saca de las sendas que no llevan a ninguna parte” (Popper 1934, p. 250).

Como es natural, tanto la postura de Popper como la de Kuhn han sido cuestionadas. Así, Lakatos ha propuesto un cambio en los planteamientos de base, defendiendo la idea de que ninguna teoría está exenta de ambigüedades y que siempre es posible cometer errores u omisiones. Para Lakatos, se trata de un proceso de “incesante mejora de las conjeturas [*guesses*], gracias a la especulación y a la crítica, siguiendo la lógica de pruebas [*proofs*] y refutaciones”¹. No obstante, ello no debiera impresionarnos puesto que “la virtud de una prueba lógica no es que nos conmine a la creencia, sino que nos sugiere dudas” (Lakatos

¹ Lakatos 1963a, p. 6. Es interesante señalar que Lakatos emplea el término “prueba” no en el sentido de demostración sino en el de mera constatación de lo afirmado por una teoría o teorema.

1963b, p. 227), llegando a establecer que “la analogía entre las ideologías políticas y las teorías científicas es (...) de mayor alcance del que normalmente es apreciado: las ideologías políticas, que inicialmente pueden ser debatidas (y quizá sólo aceptadas bajo presión), pueden convertirse en un incuestionable conocimiento de fondo incluso en una sola generación: los críticos son olvidados (y quizás ejecutados) hasta que una revolución rehabilita [*vindicate*] sus objeciones” (1963b, p. 228). De este modo, “en contra del falsacionismo ingenuo, ningún experimento, informe experimental, enunciado observacional o hipótesis falsadora de bajo nivel bien corroborada puede originar por sí mismo la falsación. No hay falsación sin la emergencia de una teoría mejor” (Lakatos 1978, p. 35). Surge así, en su propia formulación, el problema de encontrar un criterio de demarcación válido, para lo cual Lakatos propone que una teoría (un “programa de investigación” en su terminología) puede considerarse científica si es empíricamente progresiva (lo que implica serlo teóricamente), esto es, si predice hechos nuevos y explica parte de estos además de los ya conocidos: “resulta esencial, para realizar una evaluación racional, el decidir qué programa fue el primero en anticipar un hecho nuevo y cuál acomodó sólo posteriormente el hecho ya conocido” (Lakatos 1978, p. 116). La novedad —y su relación con la cuestión del papel de los datos en el progreso científico— estriba en que, de acuerdo con este planteamiento:

“Dichas conjeturas, muertas y ahora olvidadas, fueron las que sugirieron los hechos y no los hechos las conjeturas. Las conjeturas ingenuas no son conjeturas inductivas: llegamos a ellas por ensayo y error, por medio de conjeturas y refutaciones” (Lakatos 1964, p. 303); “Hoy es posible demostrar con facilidad que no se puede derivar válidamente una ley de la naturaleza a partir de un número finito de hechos, pero todavía es posible leer acerca de teorías científicas que son probadas por los hechos. ¿A qué se debe esa terca resistencia a la lógica elemental?” (Lakatos 1978, p. 2).

En este sentido, afirmaciones como la de que “la ciencia se está haciendo cada vez más dependiente de los datos, además las tecnologías para el procesamiento de datos no fueron diseñadas para la escala y heterogeneidad de los datos en el mundo moderno” (Fox y Hendler 2009, p. 147) pueden resultar muy sugerentes pero, en realidad, no aportan nada nuevo. Que el mundo se torna cada vez más complejo es algo conocido; es algo que sucede desde la revolución neolítica, por mencionar algún periodo concreto. Los experimentos de Arquímedes requerían menos datos que los de Tycho Brahe (de ahí el telescopio) y, los de éste último, menos que la secuenciación del genoma humano (y de ahí el multiprocesamiento

simétrico). No obstante, la necesidad de más datos no es capaz por sí sola de usurparle al método científico su papel como garante y guía del progreso científico. Ello supondría un error todavía más grave que aquél del que advirtiera Hayek —confiar más en la ciencia que en el método científico—, cual es el hecho de confiar más en la tecnología que en el método científico. Por ello, si también en este caso deseamos “salvaguardar la reputación de la ciencia y prevenirnos de una usurpación [*arrogation*] del conocimiento, (...) deberá dedicarse considerable esfuerzo a desacreditar tales usurpaciones” (Hayek 1974).

La clave reside pues en la consecución de un adecuado equilibrio entre observación y entendimiento, algo que, como se ha visto con cierto detalle en el apartado anterior, no constituye una apreciación en absoluto reciente. Ya Kant (1781, p. 97) advirtió:

“Sin sensibilidad ningún objeto nos sería dado y, sin entendimiento, ninguno sería pensado. Tampoco pueden ambas facultades o habilidades intercambiar sus funciones. Ni el entendimiento es capaz de intuir nada, ni los sentidos son capaces de pensar nada. Sólo de la unión de ambos puede surgir el conocimiento”.

No resulta por tanto conveniente centrar la atención sobre los datos sin más, sino que cada campo de investigación ha de abordarse de un modo orgánico. La capacidad de recoger y procesar datos es importante, pero lo es más la capacidad para interpretarlos o, si se diera el caso, desecharlos. En este contexto, y aún a riesgo de parecer un símil algo exagerado pero sin duda muy ilustrativo, podrían entenderse ciertos planteamientos de los gurús de los *Big Data* como si postularan una especie de ciencia “a martillazos” que sustituya la metodología y las técnicas de toda la vida por una suerte de “atajo tecnológico”. En cualquier caso, no se trata de desaprovechar los importantes resultados que dicha tecnología pueda ofrecer, sino simplemente de relativizar el empeño por ver en ella el advenimiento de un cambio de paradigma (el llamado *cuarto paradigma*; Hey, Tansley y Tolle 2009) o un todavía más grandilocuente “fin de la teorías”, una cuestión que se abordará más adelante.

3. Más datos no es mejor

Una vez revisado el papel de la metodología frente al de los datos en sí mismos o la idea de “Data” en los *Big Data*, es posible pasar a analizar la problemática relativa a la idea de “Big”. En otras palabras, la importancia de disponer de *muchos* datos o cuál sea el valor añadido por el mero hecho de incrementar la cantidad de éstos. En este sentido, difícilmente pueda

cuestionarse que existen ámbitos científicos concretos, como la genética, en los que el empleo masivo de datos ha abierto la puerta a importantes avances. No obstante, la cuestión reside en si ello constituye realmente un campo nuevo (llámese *Big Data*, e-ciencia o de otro modo) o si simplemente se trata de la genética de siempre apoyada por la informática de siempre. Así, el término “*Big Data*” podría estar convirtiéndose en una etiqueta más dentro de las estrategias de *marketing*, del mismo modo que lo puedan ser los términos “*Light*” o “*Bio*” en Alimentación. Surge aquí pues el peligro propio de cuando no está del todo clara la frontera entre lo que pueda ser catalogado de una estrategia de *marketing* o un mero fraude.

En lo relativo a los datos, no parece correcto afirmar que “cambiando la cantidad, se cambie la esencia” (Mayer-Schönberger y Cukier 2013, p. 10). Al menos, no parece correcta la generalización de semejante afirmación, puesto que no siempre “más” significa “distinto” (y mucho menos “mejor”). Recogiendo un símil propuesto por Taleb (2008, p. 114), “escuchar las noticias en la radio cada hora es mucho peor para uno que leer un semanario, porque el intervalo más largo permite que la información se filtre un poco”. En definitiva, más datos no significa necesariamente mejor y, en cualquier caso, el caer en semejante error sí resulta inherentemente peor.

A pesar de la actual capacidad de acceso a grandes cantidades de información, la naturaleza de los datos sigue siendo la misma y el problema de la representatividad de una muestra (por citar un aspecto) permanece, sucediendo —como es de sobra conocido— que una muestra no es más representativa *sólo* por el hecho de ser más amplia. No se trata sólo del “cuántos” sino del “cómo” y no hay duda de que un adecuado control sobre el origen y la validez de los datos (los cuales pueden ser “débiles”, “sesgados”, “irrelevantes”, “ininterpretables”, etc.) resultará más difícil conforme el tamaño de la muestra crezca. El mejor conocimiento de un fenómeno no dependerá únicamente de una mayor disponibilidad de datos, por lo que no debiera aceptarse que sea *la mera disponibilidad* de datos la que determine su valor; ésta no puede convertirse en un *criterio* de muestreo que termine por convertir los datos masivos en sesgos masivos. Por supuesto que la disponibilidad de datos puede resultar muy positiva, pero no puede hablarse de un progreso científico (una mejora en la comprensión de un fenómeno) basándose exclusivamente en la cantidad de datos que haya sido posible recoger.

Pese a lo expuesto, la esencia de lo que se ha llamado la falacia *Big Data* es la de sostener que cuanto mayor sea la cantidad de datos disponibles, mejor será la caracterización que de la realidad se haga (llámese modelo, patrón o simulación). La cantidad de datos no es equiparable con la cantidad de “verdad”, siendo aplicable aquí aquella frase de Ortega y Gasset (1914, p. 18): “¿sabemos tantas cosas que no comprendemos!”. *Big Data* se transforma así en una especie de “burbuja científica” o “fiebre del oro” en la que no se termina de comprender que para encontrar una pepita se requiere de mucho y meticoloso bateo. La información ha de recabarse en *profundidad* y no en *cantidad*. Han de buscarse *Smart Data* y no *Big Data*. Ejercerse tareas de inteligencia y no de mero acopio de datos. Así, la importancia real de los datos se limita al hecho de constituir un mero potencial de información. Pese a ello, el *furor Big Data* está evidenciando que no constituye una comprensión intuitiva la distinción entre datos e información. No toda la información almacenada en los datos masivos resultará valiosa o significativa. Además, el ritmo de crecimiento de la información extraíble de los datos no es proporcional al ritmo con el que se incrementa la cantidad éstos. Por expresarlo matemáticamente, puede que los datos crezcan exponencialmente pero la información lo hará a menudo de forma lineal. De hecho, tal y como señala Healey (2012), “la mayoría de las empresas no sobresalen en el uso de los datos que tienen ahora. ¿Cuáles son las probabilidades de que realicen mejores análisis agregando más volúmenes de información sin cambiar sus estrategias?”. Lo que nos recuerda que a menudo, tal y como expresara Wittgenstein (1960, p. 47), “los problemas no se resuelven aportando nueva información, sino poniendo en orden la que hemos conocido siempre”. En este sentido, resulta muy significativa la frase de Crawford (2013) al rechazar el mito de que *Big Data* hace “inteligentes” a las ciudades: “Big data es únicamente tan bueno como lo sean los administradores de la ciudad que están empleando y analizando estos datos”.

Frente a ello se alega que “toleramos las patologías de la cuantificación —una forma de conocimiento árido, abstracto y mecánico— porque los resultados son muy poderosos. (...) En ciencia, en los negocios y en los más razonables sectores del gobierno, los números han ganado limpiamente y sin paliativos [*fair and square*]” (Wolf 2010). Pero lo cierto es que, de acuerdo con varias de las cuestiones tratadas hasta aquí, no resulta en absoluto correcto afirmar que sean los números los “vencedores”, ni que lo sean “limpiamente y sin paliativos”. En primer lugar, ya se ha visto cómo esa hipotética “victoria” estaría más del lado de los métodos que de los números. En segundo lugar, se ha explicado cómo tales “números”, por sí

solos, difícilmente podrían hacer nada, no teniendo por tanto siquiera sentido el plantear su capacidad para hacer algo “limpiamente”, tal y como alega Wolf. Si la disputa fuese entre disponer de datos o no, la cosa estaría clara: ganan los datos; gana el estar informado. Pero las intenciones que se intuyen tras la pretensión de semejante “triunfo” parecen muy distintas. Al plantear la cuestión en términos de una competición en la que un contendiente es “la cuantificación”, resulta inevitable pensar que el otro competidor sea lo “cualitativo”, de modo que, furtivamente, se sitúa a la tecnología por encima del entendimiento humano y el método científico. Es una actitud que recuerda sobremanera a aquella advertencia de Horkheimer y Adorno (1944, p. 63) en torno a la obsesión por una “calculabilidad del mundo” a través de la cual “el número se convirtió en el canon de la Ilustración”, de manera que “todo lo que no se agota en números (...) se convierte (...) en apariencia; el positivismo moderno lo confina en la literatura”.

Desde ciertos sectores se transmite la idea de que la solución a ciertos problemas que, tras décadas de investigación “la ciencia” no ha podido resolver (cáncer, cambio climático, terrorismo, etc.), va a venir de la mano de más datos. Se afirma: “podemos arrojar los números dentro de los más grandes clusters de cálculo que el mundo haya visto y dejar que los algoritmos estadísticos encuentren patrones donde la ciencia no puede” (Anderson 2008) y se anuncia con ello “el fin de la teoría”.

“Algoritmos estadísticos (...) donde la ciencia no puede”. ¿No suena eso a una estadística “a martillazos”? Este tipo de afirmaciones son las que, a nuestro entender, deberían originar un debate epistemológico (y ético) en torno a las expectativas puestas en *Big Data*. Tal y como ya se ha dicho, puede que los datos masivos sean capaces de abrir la puerta a grandes descubrimientos, pero jamás *harán ciencia*. Jamás constituirán una alternativa al método científico, sino su complemento. Es cierto que resulta seductora la idea de poner un ordenador (o miles de ellos) a recabar información para luego procesarla y, simplemente, “ver qué sale”. Resulta más atractivo abrazar aquél viejo adagio latino según el cual “hasta durmiendo, la red pesca” (*dormientis rete trahit*), olvidando aquel otro que nos recuerda que “el camino a las estrellas discurre entre dificultades” (*per aspera ad astra*). Ello nos recuerda aquella situación que, de nuevo, ya denunciaron Horkheimer y Adorno (1944) entorno a la configuración de la modernidad: “el pensamiento pierde impulso, se limita a la aprehensión del hecho aislado. Las conexiones teóricas complejas son rechazadas como fatiga inútil y

molesta” (p. 240), lo cual sólo puede desembocar en la “autodestrucción de la Ilustración” (p. 53) como consecuencia del “triunfo de la razón tecnológica sobre la verdad” (p. 182).

Otro aspecto a tener en cuenta con relación a las capacidades inherentes a los datos es el de la llamada “maldición de la dimensión” (Bellman 1961), y que hace referencia al hecho de que nuestra capacidad de ofrecer resultados con las suficientes garantías probabilísticas (nivel de significación) queda mermada por el simple hecho de que el número de variables intervinientes en el fenómeno estudiado aumente. En este sentido, resulta sorprendente la facilidad con la que la sobreabundancia de datos produce (o, quizás, debiera producir) una peculiar mezcla entre aturdimiento y embelesamiento. De hecho, algunos autores como Galli (2013), han propuesto experimentos en los que se simula el tratamiento de datos relativos a un mismo fenómeno, por un lado en las cantidades habituales y, por otro, en las cantidades propias de los *Big Data* (mil veces mayores, pongamos por caso). En tales experimentos, y pese a tratarse de datos procedentes de algoritmos de generación de números pseudoaleatorios (y por tanto sin correlación alguna), en el caso de trabajar con datos masivos se encontraban con relativa facilidad buenas correlaciones. Como si, perdida en la abundancia de cantidades, la dificultad de encontrar patrones (en este caso un simple modelo lineal) se desvaneciera. Ello vendría a intensificar ése exceso de confianza en el papel de los datos que ya ha sido referido y que podría desembocar en cierta tendencia a la apofenia, como si se pensara: “No es que no hubiera un patrón, es que no fuimos capaces de reconocerlo. Todo era una cuestión de datos”. De ser así, ahora que tales datos están por fin al alcance, es posible que veamos proliferar patrones, modelos y correlaciones por todas partes. En la actualidad ya es posible encontrar investigaciones que utilizan los datos masivos de las redes sociales para el estudio de los fenómenos más variopintos; desde huracanes o terremotos (Sakaki, Okazaki y Matsuo 2010) a incendios forestales (Goodchild y Glennon 2010). En este sentido, las conclusiones de Galli (2013) son crudas pero posiblemente, bien encaminadas:

“En estos casos sólo el sentido común —educado— puede descartar correlaciones casuales. Aún peor, cuando hay mucha cantidad de datos es muy difícil separar el ruido de la señal. Cada día veremos más gráficos y análisis de este tipo, totalmente inservibles y que en algunos casos podrían llevar a generar medidas o políticas desastrosas si no se analizan con cuidado”.

4. Big data. ¿Muerte de la causalidad?

Tal y como se acaba de ver, el enaltecimiento de los datos, que con la “revolución” *Big Data* se arrojan la capacidad de hacer ciencia por sí solos, ha conducido de un modo casi directo a una relativización de la importancia para el progreso científico de la enunciación de teorías que permitan explicar la realidad de un modo general. Para hacer esto posible, uno de los obstáculos que se ha eludido ha sido el del concepto de causalidad. Hasta ahora se habían podido apreciar dos actitudes fundamentales frente al principio de causalidad: aceptarlo o rechazarlo. Lo que vienen a proponer las técnicas *Big Data* es ignorarlo, lo que, en la práctica, equivale a considerarlo intrascendente, algo difícil de admitir. Se entiende que si se disponen de tantos datos relativos a un suceso, encontrar su causa (explicar) es lo de menos. Así, se ha sugerido que grandes cantidades de datos podrían poseer el potencial necesario para “apartarnos de la antigüedad [*age-old*] de la búsqueda de la causalidad” (Mayer-Schönberger y Cukier 2013, p. 14) permitiendo sustituir la exactitud de la causalidad por unos niveles de correlación “suficientemente buenos” (pp. 14, 52 y 195). Así, se admite que, en realidad, “big data trata sobre el *qué*, no sobre el *por qué*” (p. 14), algo que casi podría interpretarse como “*Big Data* encuentra algo, pero no sabe el qué” (dado que no lo *explica*).

En este sentido, podría resultar legítimo alegar que esta sustitución de la verdad por la probabilidad supone abrazar una cierta forma de escepticismo, dado que con ello el investigador limita de antemano sus propias posibilidades y las de la ciencia misma, asumiendo una imposibilidad de alcanzar la meta que él mismo se habría propuesto desde los albores de la civilización: el conocimiento preciso de la realidad que le rodea. Una “capitulación de la razón ante la realidad” o “autoliquidación del pensamiento e instrumentalización de la verdad” (Sánchez 2010, p. 24) por la cual, y citando a Horkheimer (1947), “el pragmatismo justificó desde sus comienzos implícitamente la difundida sustitución de la lógica de la verdad por la de la probabilidad, que desde entonces es la que ha venido, con gran diferencia, a convertirse en la predominante” (p. 76), de manera que “la probabilidad, o, mejor aún, la calculabilidad sustituye a la verdad” (p. 77).

Ciertamente, en ocasiones esa comprensión del mundo en clave *Big Data* sólo parece entenderse en el marco de aquel desplazamiento hacia un pensamiento “débil” que denunciara Vattimo (1988) y por el que se renuncia a la pretensión de reconstruir los hechos con vistas a

encontrarles un sentido o, en otras palabras, descubrir una causalidad. Una actitud que podría perfectamente preludear esa ya referida introducción de la ideología en el ámbito científico que llevó a Lyotard (1991, p. 14) a afirmar que “el saber científico es una clase de discurso”, algo producido para ser vendido y consumido y que, al dejar de constituir su propio fin, queda seriamente devaluado.

Big Data acomete por tanto otro de los problemas clásicos de la filosofía de la ciencia. De acuerdo con el principio de causalidad, el ser humano ofrece la explicación a un hecho siempre que es capaz de explicitar las causas que lo han producido. En otras palabras, sería gracias al principio de causalidad que el ser humano *se* explica multitud de fenómenos, algo que, por otro lado, Hume consideró absurdo e inalcanzable ya a finales del siglo XVII. Para Hume, por más veces que se verifique la secuencialidad de dos hechos concretos, jamás se estará, siendo rigurosos, en condiciones de afirmar que exista una causalidad general. Como mucho será posible formular una hipótesis al respecto. De este modo, Hume no encontraría otra base epistemológica para la causalidad que el mero hábito o costumbre; nos acostumbramos a que tras determinado hecho se produzca otro y a esa mera *sensación* —de seguridad, por otra parte— es a lo que nosotros llamamos “causa”. Frente a tal escepticismo, la postura opuesta sería la de considerar que las verdades *de hecho* son verdades *de razón*, siendo sólo nuestra incapacidad de acceder a un conocimiento *completo* —que tampoco es capaz de facilitarnos *Big Data*— el único motivo por el que en determinadas ocasiones no apreciamos la razón que se oculta detrás de ciertos sucesos. Con relación a la necesidad de una idea de causalidad, ya en el siglo XIX Venn citaría a De Morgan para sugerir que “«los [mismos] fundamentos de la teoría de la probabilidad han dejado de existir en la mente que ha formado la idea», «de que haya sucedido o vaya a suceder nada sin alguna razón concreta por la cual deba haber sido exactamente como fue y no de otro modo»“ (Venn 1866, p. 327).

Se aprecia, pues, la centralidad del concepto de causalidad en la elaboración de teorías dentro de cualquier campo científico. Para algunos autores, este impulso aparentemente irrefrenable por la identificación de una causalidad en los fenómenos que contemplamos, podría tener su origen en lo que Taleb (2008, p. 63) ha denominado la “falacia narrativa” y que, no sin cierta razón, ha calificado de auténtico fraude. Para Taleb, “la narratividad nace de una arraigada necesidad biológica de reducir la dimensionalidad [de la información que nos llega]” que llega a entrañar que “no teorizar *es un acto*”, siendo el teorizar la “opción «por

defecto»“ (p. 64). En opinión de Taleb, “la planificación está asociada al paquete que nos hace humanos, a saber, a nuestra conciencia” (p. 64), constituyendo en definitiva “una patología de la naturaleza humana” (p. 173). Un planteamiento que recuerda a aquella “enfermedad que afecta a la razón” de la que ya hablara Horkheimer (1947) y que, en su opinión, no sería sino un mero “afán de dominio” sin un origen “en un momento histórico determinado, sino como inseparable de la esencia de la razón en la civilización” (p. 179).

Sea como fuere, resulta ciertamente difícil negar esta tendencia del ser humano hacia la narratividad, así como la posibilidad de que dé origen a una cierta sobrevaloración de los fenómenos de causa-efecto que tanto representan para nuestro progreso científico. También resulta difícil no reconocer que ello supone, en ocasiones, desvirtuar la realidad estudiada, al simplificar —para hacerlos comprensibles— determinados fenómenos que, o bien resultan excesivamente complejos —lo que la teoría algorítmica de la información denomina complejidad de Kolmogorov— o bien, sencillamente, son impredecibles (tienen un carácter estrictamente aleatorio). Podrá interpretarse tal tendencia como un deplorable rasgo de arrogancia o como un simple mecanismo de supervivencia, pero lo cierto es que está ahí. Es nuestra particular lucha contra ese *horror vacui* que nos produce todo lo arbitrario e inabordablemente complejo y en cuyo contexto Taleb ha comparado el papel desempeñado por las modernas teorías científicas con el de los mitos en la antigüedad. En cualquier caso, no parece razonable ese tránsito radical por el que deba pasarse del fideísmo más ingenuo —aquel “pensamiento infantil del hombre” del que hablara Carlyle (1841, p. 46)— al más desmedido cientifismo. En el caso que nos ocupa, se trata simplemente de encontrar la interpretación adecuada al papel que deba desempeñar la adquisición de datos en la futura investigación científica. Frente a las expectativas desatadas por los datos masivos, entendidos como la llave tecnológica que abrirá la puerta a incontables secretos, resulta necesario que los abanderados de los *Big Data* expongan y demuestren con datos —también para esto— que dicha caracterización es correcta. Se ha de traducir a hechos lo que hasta ahora apenas pasa de ser una mera retórica *Big Data*. Es un reto que sus defensores deben abordar y que ha sido inventariado como sigue (Boyd y Crawford 2012):

1. “*Big Data* cambia la definición de conocimiento”.
2. “Las pretensiones de objetividad y exactitud son engañosas”.
3. “Más datos no siempre implica mejores datos”.

4. “Fuera de un contexto, *Big Data* pierde su significado”.
5. “Sólo porque sea accesible no lo hace ético”.
6. “El acceso restringido a *Big Data* genera nuevas brechas digitales”.

En lugar de esto, la mayoría de trabajos se limitan a describir las técnicas empleadas (apenas novedosas en el fondo) y qué decisiones fueron tomadas, olvidando demostrar que el resultado final fue *causado* por tales decisiones (algo que sí validaría sus postulados). Por citar un caso muy conocido, ésta es la situación con relación a la campaña de Barack Obama en el año 2012, durante las elecciones presidenciales en Estados Unidos. Se expone cómo Obama ganó las elecciones gracias, supuestamente, al empleo de técnicas *Big Data*, pero lo cierto es que las características del fenómeno, así como las pequeñas variaciones alegadas en los porcentajes de votos no permiten, siendo honestos, hablar de factores *decisivos* en la victoria de Obama. Resulta imposible cualquier análisis que pretenda comparar los resultados de una campaña electoral en un escenario “libre de datos masivos” frente a otro en el que éstos estuvieran presentes, algo que permitiría evaluar la influencia *efectiva* de tales técnicas. No es posible hablar, como hacen algunos medios, de un vuelco electoral “dirigido”, sino simplemente de individuos que reflexionan y votan, normalmente en base a sus preferencias y no a aquellas que les suponga un grupo de expertos en *data crunching*.

En este contexto, surge el problema de que las expectativas generadas por el empleo de datos masivos, podrían reforzar la presencia de una cierta forzosidad a la hora de elaborar modelos que ya no basta con que sean descriptivos sino que han de ser predictivos. Tal y como dice Taleb (2008), “la predicción está firmemente institucionalizada en nuestro mundo” (p. 153), pero lo cierto es que la naturaleza de determinados fenómenos, como puedan ser el comportamiento humano o el de la atmósfera, son de naturaleza caótica o no lineal. Así, sucesos en apariencia irrelevantes y completamente imprevisibles son capaces de alterar por completo la dinámica sometida a estudio, por lo que debiera admitirse como algo normal el rechazo a realizar cierto tipo de predicciones sin que ello deba suponer pérdida de autoridad alguna. Se trata de admitir que el “medio” por el que discurren determinados fenómenos es ése áspero y desapacible “Extremistán” del que habla Taleb en *El cisne negro*, a pesar de que se tienda a suponerlos aconteciendo en un amable y complaciente “Mediocristán”. Se ignora casi de manera sistemática una eventual incapacidad de abarcarlo todo, lo que nos lleva a la necesidad de evaluar cuestiones como la tesis de Duhem-Quine (Quine 1951) o el llamado

holismo epistemológico. De acuerdo con esta tesis, la complejidad lógica de cualquier teoría es tal que implica infinitas potenciales situaciones empíricas a explicar a partir de unos datos que, por muy numerosos que sean, siempre serán finitos. En consecuencia, toda teoría siempre queda subdeterminada por los datos (algo que se conoce como infradeterminación empírica de las teorías) resultando imposible la realización de experimentos que puedan considerarse auténticamente definitivos.

Lo cierto es que el hecho de no modelizar no constituye un problema tan grave. Es posible realizar importantes aportaciones sin recurrir a una modelización que se reivindica como una aproximación válida a la realidad pero que en ocasiones resulta excesivamente simplificadora. Hay situaciones en las que nuestra mejor aproximación puede no ser otra cosa que una burda adulteración —aquél “supongamos que la vaca es esférica” (Harte 1988)—. En tales casos sería preferible abstenerse de predecir pero no así de “decir” (esto es, describir). Es perfectamente posible planificar incluso cuando no sea posible planificar perfectamente. Conocer *sin teorizar* puede resultar de gran utilidad dado que, al fin y al cabo, como ironiza Taleb (2008), “una teoría es como un medicamento (o un gobierno): a menudo inútil, a veces necesaria, siempre interesada y, en ocasiones, letal” (p. 285).

5. Big data. ¿Un cambio de paradigma?

El término *Big Data* no es del todo reciente, sino que surge como problema (no como solución) hacia finales de los noventa (Cox y Ellsworth 1997, p. 235). Pese a ello, en los últimos años se ha pretendido ver en ellos una revolución, no de carácter tecnológico, sino científico y casi podría decirse que filosófico o epistemológico. Se ha llegado a hablar de la “obsolescencia del método científico” (Anderson 2008), apreciándose cómo para el fundamentalismo *Big Data*, estaríamos entrando en un escenario en el que la mera *datificación* del mundo posibilitará la revelación de toda su complejidad. Una actitud que hace pensar más en una “deificación” del dato que en una *datificación* sin más, puesto que de ella surgen afirmaciones como que con *Big Data* se abre un “*impasse* histórico en el que «Dios ha muerto». Es decir, las certezas en las que creíamos, de nuevo están cambiando” (Mayer-Schönberger y Cukier 2013, p. 18). Tal y como ya se ha dicho, se pretende comprender un fenómeno por el mero hecho de disponer de muchos datos acerca de él y se

presenta como revolución lo que no parece sino una reedición más del viejo positivismo de siempre, si acaso con ropajes más *geek*.

En este sentido, lo que plantean algunos defensores de los *big data* supone aquello que Popper denominaba una “hipótesis audaz”, esto es, una propuesta seductora que promete mucho pero queda expuesta al hecho de que, cuanto más promete una teoría (cuanto más pretende abarcar), menor es su grado de falsabilidad y, en consecuencia, mayor es el riesgo que se corre al aceptarla. Por ello, y dado que, como dijera Heidegger (1947, p. 82), “quien piensa a lo grande, ha de errar a lo grande”, a mayores promesas, mayores controles. Especialmente cuando un triunfo de la irrupción de los datos masivos en nuestras vidas podría tener consecuencias imprevisibles no sólo a nivel científico sino también a nivel social — afectando para empezar a nuestra privacidad—.

Se trata en definitiva de situar la importancia de los datos en su justo lugar. El valor no está en los datos sino en lo que se haga con ellos. Si Marx (1867, p. 74) previno en economía del “carácter místico” o del “fetichismo” de la mercancía, con una producción reducida a “una inmensa acumulación de mercancías” (Marx 1867, p. 39), hoy podría hablarse en ciencia del fetichismo de los datos. Se presume que si se consiguen datos “perfectos” (que es lo que en el fondo muchos entienden cuando hablan de datos “masivos”) se accederá a un conocimiento “perfecto”, ya sea del comportamiento humano, la naturaleza o lo que sea el caso. En ocasiones, casi se intuye el mensaje “no hace falta más ciencia, hacen falta más datos”. Pero lo cierto es que la acumulación indiscriminada de datos no sólo podría no resultar de ayuda, sino que podría convertirse en una carga, como la de aquel protagonista de *Funes el memorioso* de Borges, incapaz de olvidar y, con ello, de pensar, puesto que “pensar es olvidar diferencias, es generalizar, abstraer. En el abarrotado mundo de Funes no había sino detalles, casi inmediatos” (Borges 1944, p. 130).

Se ha caracterizado la llegada de los *big data* como un “cambio de paradigma”, hablándose de un “cuarto paradigma” (Hey, Tansley y Tolle 2009) en lo relativo a la forma de hacer ciencia. Se afirma que las técnicas tradicionales no sirven para el tratamiento de cantidades tan extraordinarias de datos. En este sentido, es posible que las técnicas tradicionales no sean capaces de procesar tanta información de un modo eficiente —aquella “aguja” que se buscaba, se pierde todavía más conforme se amplía el “pajar”—, pero se

confunde tecnología con ciencia y, por muy conectadas que ambas estén, no son lo mismo. No hay, en definitiva, tal cambio de paradigma *científico*.

Al revisar esta comprensión de los datos masivos como un “cuarto paradigma” “consistente en las técnicas y tecnologías necesarias para llevar a cabo ciencia intensiva en datos [*data-intensive science*]” (Bell, Hey y Szalay 2009, p. 1297), así como la “transformación” (Hey, Tansley y Tolle 2009, xvii-xxx) que haya podido suponer en el método científico tradicional, Wilbanks (2009, p. 210) habla de la necesidad de que la ciencia se adapte a un nuevo escenario de “aluvión de datos [*data deluge*]” que sitúa a los datos en “el corazón de este nuevo paradigma”. No obstante, el propio Wilbanks relativiza la idea de un cambio de paradigma admitiendo que no se trata “realmente de un cambio de paradigma en el sentido dado por Kuhn [*in the Kuhnian sense*]” sino que “lo que necesitamos es cambiar nuestro paradigma de nosotros mismos como científicos —no los viejos paradigmas del descubrimiento”. De este modo, no parece necesario, tal y como se ha afirmado, “plantearnos una transformación en el marco conceptual epistemológico” (Vallverdú 2008, p. 209). El marco epistemológico es el de siempre, la ciencia no. Ciencia es saber objetivo y verificable, por lo que un cambio epistemológico en el paradigma *científico*, etimológicamente hablando, implicaría un cambio en el procedimiento seguido para acceder a dicho saber y verificarlo, una tarea para la cual, siendo rigurosos, los celebrados *Big Data* no pueden aportar nada nuevo o, al menos no de forma tan sensacionalista.

La principal característica que define a los datos masivos no es la de trabajar con datos sino la de hacerlo con una cantidad extremadamente grande de éstos. Ello supone, evidentemente, que alguna de las técnicas a emplear sea revisada pero, bien mirado, no hay mucho más. Desde el advenimiento de las técnicas de generación de números pseudoaleatorios, y en especial del método Montecarlo, la estadística viene trabajando con cantidades ingentes de datos, resultando aquí en cierto modo irrelevante el que sean de carácter real o simulado. La estadística tradicional no pierde su validez por el hecho de manejar tantos datos, como la minería no deja de serlo por ser aplicada a una veta enorme. De hecho, ya se ha hecho mención a la falta de un criterio claro que establezca qué pueda recogerse bajo el sello “*Big Data*” y qué no. Por otro lado, puede afirmarse que un planteamiento es radicalmente nuevo cuando permite la resolución de problemas clásicos mediante la adopción de un planteamiento que no se había concebido ni aplicado con

anterioridad y, si se pretende caracterizar como un nuevo paradigma, tal cambio habría de producirse en la forma de cierta ruptura con lo anterior. Sin “falsación” del paradigma anterior, no puede aceptarse otro nuevo, por muchas y muy seductoras que sean sus promesas. En este sentido, *Big Data* no constituye ningún nuevo paradigma por la sencilla razón de que, como planteamiento, ni siquiera es genuinamente nuevo. Difícilmente los analistas de datos podrán progresar al margen del paradigma clásico, tomándolo como algo obsoleto ni, menos aún, podrán avanzar a través de un caótico escenario “libre de teorías” o transformado en eso que Horgan ha denominado “ciencia irónica” (Horgan 1996, pp. xxxi y 24). Tal y como observa Healey (2012), “atrassa el reloj unos pocos años y substituye las palabras «ingeniero de virtualización» o «programador Cobol» o incluso «webmaster» por «especialista en big data» y de modo semejante encontrarás personas capaces de predecir catástrofes”.

Big Data ha irrumpido en la vida científica con cierta estridencia, algo que aquí se trata aquí de evaluar si está justificado. No tiene sentido hacer de los datos masivos un hito revolucionario, un cambio en el paradigma científico o, peor aún, un cambio en nuestra cosmovisión como comunidad científica. Semejante actitud podría tener su origen en ese rasgo tan acostumbrado hoy que es el de confundir el impacto mediático con la verdadera relevancia de un acontecimiento (sea científico o no), algo que Taleb ha reconocido al afirmar que “los científicos también somos vulnerables a las narraciones, poniendo el énfasis en los remates y títulos (...) «sexys» que llamen la atención antes que en cuestiones más sustanciales” (Taleb 2008, p. 75). Sin embargo, en ocasiones lo verdaderamente revolucionario, surge en la forma de un discreto logro que, tras un periodo inicial durante el cual pasa casi desapercibido, abre con el tiempo líneas de investigación que pasan a constituir los pilares fundamentales de logros posteriores. Ante tal realidad, los científicos pueden callar, permitiendo el libre desarrollo de lo que podríamos denominar una ciencia “viral”, erigida sobre la dictadura de lo impactante y lo anecdótico, o pronunciarse y seguir haciendo, simplemente, ciencia.

En definitiva, resulta completamente natural que, como promesa tecnológica, las técnicas *datalíticas* sean sometidas a un estricto control por parte de la sociedad en general y de la comunidad científica en particular. Podría citarse en este sentido aquello de lo que advirtió Heidegger (1954, p. 13) cuando analizara nuestra relación con la técnica:

“Nunca experimentamos nuestra relación con la esencia de la técnica mientras nos limitemos únicamente a representar y practicar lo técnico, mientras nos conformemos con lo técnico o lo evitemos. En todas partes nos mantenemos dependientes y encadenados a la técnica, tanto si la afirmamos apasionadamente como si la negamos. Sin embargo, lo más ingenuo es que estemos abandonados a la técnica, si la consideramos como algo neutral; ya que esta concepción, a la que hoy se rinde especial pleitesía, nos hace completamente ciegos frente a la esencia de la técnica”.

Por ello quizás debiera evitarse ese terco empeño por hablar de “el fin de...” (la Física, la Filosofía, etc.) que revisara Horgan (1996) en *El fin de la ciencia*. Con el surgimiento de los datos masivos no se produce, afortunadamente, el fin de nada, mucho menos el fin de la ciencia o de nuestro paradigma científico. Se trata del alumbramiento de una oportunidad que, bien asistido, abrirá puertas interesantes pero que, mal entendido, sólo puede extraviarnos o generar problemas en lugar de resolverlos.

Referencias bibliográficas

- Alpert, M. y Raiffa, H. (1982): “A progress report on the training of probability assessors”, en *Judgment under uncertainty: Heuristics and biases*, eds. Daniel Kahneman, Paul Slovic y Amos Tversky, pp. 294-305, Cambridge, Cambridge University Press.
- Anderson, C. (2008): “The End of Theory: The Data Deluge Makes the Scientific Method Obsolete”, *Wired Magazine* (16 julio de 2008). http://archive.wired.com/science/discoveries/magazine/16-07/pb_theory (acceso 1 de noviembre de 2015).
- Bell, G., Hey, A.J.G. y Szalay, A. (2009): “Beyond the Data Deluge”, *Science*, vol. 323 (5919), pp. 1297-1298.
- Bellman, R.E. (1961): *Adaptive control processes: A guided tour*, Princeton, Princeton University Press.
- Berry, D.M. (2011): “The computational turn: thinking about the digital humanities”, *Culture Machine. The Digital Humanities: Beyond Computing*, 12, pp. 1-22. <http://www.culturemachine.net/index.php/cm/article/view/440/470> (acceso 1 de noviembre de 2015).
- Borges Acevedo, J.F.I.L. (1944): “Funes el memorioso”, en *Ficciones*, Madrid, Alianza [1970].
- Bowker, G.C. (2005): *Memory practices in the sciences*, Cambridge, MIT Press.

- Boyd, D. y Crawford, K. (2012): “Critical Questions For Big Data. Provocations for a cultural, technological, and scholarly phenomenon”, *Information, Communication & Society*, 15 (5), pp. 662-679.
- Carlyle, T. (1841): *Los heroes*, Madrid, Sarpe [1985].
- Cox, M., Ellsworth, D. (1997): “Application-controlled demand paging for out-of-core visualization”, en *Proceedings of the 8th conference on visualization '97*, Los Alamitos, IEEE.
- Crawford, K. (2013): “The hidden biases in big data”, *Harvard Business Review* (1 de abril de 2013). <https://hbr.org/2013/04/the-hidden-biases-in-big-data> (acceso 1 de noviembre de 2015).
- Crawford, K. (2013): “The Raw and the Cooked: The Mythologies of Big Data”, Conferencia en la UC Berkeley School of Information, Berkeley, California (30 de mayo de 2013). <http://dataedge.ischool.berkeley.edu/2013/schedule/raw-and-cooked-mythologies-big-data> (acceso 1 de noviembre de 2015).
- Descartes, R. (1637): *Discurso del método*, Madrid, Espasa-Calpe [1990].
- Fox, P. y Hendler, J. (2009): “Semantic eScience: Encoding Meaning in Next-Generation Digitally Enhanced Science”, en Hey, A.J.G., Tansley, S. y Tolle, K.M. (eds.), *The Fourth Paradigm: Data-Intensive Scientific Discovery*, Redmond, Microsoft External Research. <http://research.microsoft.com/en-us/collaboration/fourthparadigm/> (acceso 1 de noviembre de 2015).
- Galli Granada, R.A. (2012): “Sé cuidadoso con el «Big Data»”, *Ricardo Galli, de software ~ De software libre, internet, legales*, 29 de mayo de 2012 (acceso 1 de noviembre de 2015). <https://gallir.wordpress.com/2013/05/29/se-cuidadoso-con-el-big-data/>
- Gitelman, L. (2013): *«Raw Data» is an Oxymoron*, Cambridge, MIT Press.
- Goodchild, M.F. y Glennon, J.A. (2010): “Crowdsourcing geographic information for disaster response: A research frontier”, *International Journal of Digital Earth*, 3 (3), pp. 231-241.
- Harte, J. (1988): *Consider a Spherical Cow: A Course in Environmental Problem Solving*, Mill Valley, University Science Books.
- Hayek, F.A. von (1974): “La Pretensión del Conocimiento”, Conferencia en homenaje de Alfred Nobel, pronunciada el 11 de diciembre de 1974, en *Los Premios Nobel de Economía 1969-1977*, 245-258, México, Fondo de Cultura Económica [1978].

- Healey, M. (2012): “6 Lies About Big Data”, *InformationWeek* (24 de octubre de 2012). http://www.informationweek.com/software/information-management/6-lies-about-big-data/d/d-id/1107025?page_number=1 (acceso 1 de noviembre de 2015).
- Hegel, G.W.F. (1807): *Fenomenología del espíritu*, Madrid, Universidad Autónoma de Madrid/Abada Editores [2010].
- Heidegger, M. (1954): “Die Frage nach der Technik”, en *Vorträge und Aufsätze*, Pfullingen, Günther Neske.
- Heidegger, M. (1926): *Lógica. La pregunta por la verdad*, Madrid, Alianza [2015].
- Heidegger, M. (1947): “Aus der Erfahrung des Denkens”, en *Gesamtausgabe 13: Aus der Erfahrung des Denkens (1910–1976)*, Frankfurt am Main, Vittorio Klostermann [1983].
- Hey, A.J.G., Tansley, S. y Tolle, K.M. (eds.) (2009): *The Fourth Paradigm: Data-Intensive Scientific Discovery*, Redmond, Microsoft Research. <http://research.microsoft.com/en-us/collaboration/fourthparadigm/> (acceso 1 de noviembre de 2015).
- Hey, A.J.G., Tansley, S. y Tolle, K.M. (2009): “Jim Gray on eScience: A Transformed Scientific Method”, en Hey, A.J.G., Tansley, S. y Tolle, K.M. (eds.), *The Fourth Paradigm: Data-Intensive Scientific Discovery*, Redmond, Microsoft External Research. <http://research.microsoft.com/en-us/collaboration/fourthparadigm/> (acceso 1 de noviembre de 2015).
- Horgan, J. (1996): *The End Of Science: Facing The Limits Of Knowledge In The Twilight Of The Scientific Age*, Reading, Addison Wesley.
- Horkheimer, M. y Adorno, T.L.W. (1944): *Dialéctica de la Ilustración*, Madrid, Trotta [2009].
- Horkheimer, M. (1947): *Crítica de la razón instrumental*, Madrid, Trotta [2010].
- Husserl, E.G.A. (1901): *Logische Untersuchungen. Zweiter Theil. Untersuchungen zur Phänomenologie und Theorie der Erkenntnis*, Halle (Saale), Max Niemeyer.
- Kant, I. (1774): “Reflexionen zur Moralphilosophie”, en *Kants gesammelte Schriften (Akademie-Ausgabe)*, 17, Berlin, W. de Gruyter [1926].
- Kant, I. (1781): “Kritik der reinen Vernunft”, en *Immanuel Kant: Werke in zwölf Bänden*, ed. Wilhelm Weischedel, Frankfurt am Main, Suhrkamp [1977].
- Kuhn, T.S. (1962): *La estructura de las revoluciones científicas*, México, Fondo de Cultura Económica [2006].

- Lakatos, I. (1963a): “Proofs and Refutations (I)”, *The British Journal for the Philosophy of Science*, 14 (53), pp. 1-25.
- Lakatos, I. (1963b): “Proofs and Refutations (III)”, *The British Journal for the Philosophy of Science*, 14 (55), pp. 221-245.
- Lakatos, I. (1964): “Proofs and Refutations (IV)”, *The British Journal for the Philosophy of Science*, 14 (56), pp. 296-342.
- Lakatos, I. (1978): *The methodology of scientific research programmes*, Cambridge, Cambridge University Press.
- Lyotard, J.F. (1991): *La condición postmoderna. Informe sobre el saber*, Madrid, Cátedra.
- Marx, K. (1867): *El capital. Crítica de la Economía Política, vol. 1*, Madrid, EDAF, [1970].
- Mayer-Schönberger, V. y Cukier, K. (2013): *Big data: A revolution that will transform how we live, work, and think*, New York, Houghton Mifflin Harcourt.
- Ortega y Gasset, J. (1914): *Meditaciones del Quijote*, Madrid, Revista de Occidente [1975].
- Poincaré, H. (1902): *La science et l'hypothèse*, Paris, La Bohème [1992].
- Popper, K.R. (1934): *La lógica de la investigación científica*, Madrid, Tecnos [2004].
- Popper, K.R. (1963): *Conjeturas y refutaciones. El desarrollo del conocimiento científico*, Barcelona, Paidós [1991].
- Quine, W.V.O. (1951): “Two Dogmas of Empiricism”, *The Philosophical Review*, 60 (1), pp. 20-43.
- Sakaki, T., Okazaki, M. y Matsuo, Y. (2010): “Earthquake shakes Twitter users: Real-time event detection by social sensors”, en *Proceedings of the 19th International Conference on World Wide Web*, Raleigh, ACM New York, pp. 851-860.
- Sánchez Bernal, J.J. (2010): Introducción a *Crítica de la razón instrumental*, de Max Horkheimer: “Quebrar la lógica del dominio. Actualidad de la crítica de Horkheimer a la razón”, Madrid, Trotta.
- Taleb, N.N. (2008): *The Black Swan. The Impact of the Highly Improbable*, New York, Random House.
- Vallverdú Segura, J. (2008): “Apuntes epistemológicos a la e-ciencia”, *Revista de Filosofía*, 64, pp. 193-214.
- Vattimo, G. y Rovatti, P.A. (1988): *El pensamiento débil*, Madrid, Cátedra.

- Venn, J. (1866): *The Logic of Chance: An Essay on the Foundations and Province of the Theory of Probability, with Especial Reference to Its Application to Moral and Social Science*, London/Cambridge, Macmillan and Co.
- Wilbanks, J. (2009): “I Have Seen the Paradigm Shift, and It Is Us”, en Hey, A.J.G., Tansley, S. y Tolle, K.M. (eds.), *The Fourth Paradigm: Data-Intensive Scientific Discovery*, Redmond, Microsoft External Research. <http://research.microsoft.com/en-us/collaboration/fourthparadigm/> (acceso 1 de noviembre de 2015).
- Wittgenstein, L. (1960): *Philosophical Investigations*, New York, MacMillan [1968].
- Wolf, G. (2010): “The Data-Driven Life”, *The New York Times Magazine* (28 de abril de 2010). http://www.nytimes.com/2010/05/02/magazine/02self-measurement-t.html?_r=0. (acceso 1 de noviembre de 2015).