

ÍNDICE

Introducción.....	9
-------------------	---

PRIMERA PARTE EL DEBATE EN LAS CIENCIAS SOCIALES SOBRE EL BIG DATA

Capítulo 1. La polémica en las ciencias sociales acerca de la digitalización y los datos masivos.....	15
Capítulo 2. ¿Qué se entiende por big data?.....	29
Capítulo 3. La producción de los datos masivos.....	35
Capítulo 4. Una aproximación a la caracterización de los datos masivos.....	43

SEGUNDA PARTE UNA CARACTERIZACIÓN DE LAS HUELLAS Y LOS DATOS DIGITALES

Capítulo 5. Los nuevos datos digitales.....	61
5.1. El Big data como un caso particular del tradicional registro de casos.....	68
Capítulo 6. una propuesta de caracterización de las huellas digitales.....	73
Capítulo 7. de las huellas digitales a los datos digitales.....	97
7.1. Entre la lógica inscricional y la dimensión semántica. Los registros.....	107
7.2. Entre la lógica inscricional y la dimensión pragmática. Las correlaciones.....	111
7.3. De las huellas a los datos. Los procesos de producción y transformación de las huellas y de los datos. Los algoritmos.....	113
7.4. El análisis y el tratamiento de las huellas digitales. Las visualizaciones.....	121

7.4.1. La forma mapa como forma canónica de representación de las redes complejas.....	144
7.4.1.1. Una reflexión sobre los grafos y los mapas.....	153

TERCERA PARTE

UNA APROXIMACIÓN A LAS FORMAS DE VISUALIZACIÓN Y DE REPRESENTACIÓN DE LAS REDES. LA TOPOLOGÍA

Capítulo 8. Las formas de visualización de las redes y de los mapas.....	165
8.1. El escalamiento multidimensional.....	166
8.2. El análisis factorial de correspondencias múltiples.....	175
Capítulo 9. La visualización de las redes complejas.....	187
9.1. La lectura e interpretación de los mapas de las redes complejas..	197
Capítulo 10. Una propuesta de desarrollo de métodos cuali-cuanti en el big data.....	223
Capítulo 11. Una propuesta de articulación topológica de distintas prácticas y técnicas de investigación social. La topología como lenguaje de posible articulación de las investigaciones sociales convencionales y de las del big data.....	237
11.1. Algunas experiencias de aproximación topológica.....	243

CUARTA PARTE

A MODO DE CONCLUSIÓN. LAS CONCEPCIONES DE FONDO DOMINANTES EN EL BIG DATA Y LAS SOCIEDADES DE CONTROL

Capítulo 12. Las concepciones de fondo del abordaje dominante del big data. Una crítica a tres posiciones fundantes en el tratamiento del big data.....	259
12.1. El sueño de la objetividad y de la eliminación de la subjetividad.....	259
12.2. La ausencia de teoría. Su construcción inmanente a partir de los datos.....	263
12.3. Lo predictivo y lo proyectivo.....	273
Capítulo 13. A modo de epílogo. Más allá de la investigación social. Las sociedades de control.....	287

BIBLIOGRAFÍA.....	301
ÍNDICE ONOMÁSTICO.....	317
ÍNDICE ANALÍTICO.....	323
ÍNDICE DE LAS FIGURAS Y LOS GRÁFICOS.....	327

CAPÍTULO 2

¿QUÉ SE ENTIENDE POR BIG DATA?

El término *datos masivos* o *big data* parece que se usó por primera vez por Erik Larson en 1989 en un artículo de la revista Harpers y alcanzó su popularidad en 2010-2011 a partir de la publicación de tres informes de la consultora McKinsey a este respecto. Actualmente es uno de los términos más de moda en el mundo de la investigación, en los debates científicos y en los medios de comunicación más en general (el 6 de julio de 2021 el buscador de Google apuntaba la existencia de 8000 millones de menciones asociadas a dicha expresión).

Más allá de su asociación más estricta con los *datos masivos* o con los *macrodatos*, el uso del término *Big data*, como señala Vayre (2014), lo caracteriza como un sistema de producción y de cálculo, como un sistema sociotécnico que integra varias dimensiones como la masificación de los datos a la que alude primariamente la expresión inglesa, su gran *volumen*, su gran *variedad*, así como la enorme *velocidad* de su tratamiento. De ahí que se haya popularizado la caracterización del Big data a partir de las tres «V» utilizadas por primera vez en el 2001 por el analista Doug Laney (Boullier, 2015; Ollion y Boelaert, 2015; Tascón y Coullaut, 2016; Beaude y Nova, 2016; Robles et al., 2020) al punto de que dicha caracterización es utilizada como definición de Big data por la Fundación del Español Urgente, «denominación con la que se alude a un conjunto de datos tal que por su volumen, variedad y por la velocidad a la que necesitan ser procesados supera las capacidades de los sistemas informáticos habituales».

Conjunto de características más definatorias y consensuales del Big data que son enriquecidas por autores como Kitchin (2014) con otras propiedades como la *exhaustividad* en el sentido de aspirar a capturar a la totalidad de las poblaciones de forma que $N = \text{Todos}$, la *granularidad* desde la perspectiva del alcance de sus datos, de la *flexibilidad*, de la *extensionalidad* en el sentido de que se pueden agregar nuevos campos fácilmente, y de la *escalabilidad* en el sentido de poder expandirse en tamaño muy rápidamente desde la perspectiva de su producción.

En este marco y aunque en el entorno del Big data se tiende a mezclar la masificación de los datos con lo digital quizás convenga, al menos en un primer momento, diferenciarlos. En este sentido, en las primeras obras, en los primeros textos consagrados al Big data, en los textos que pueden considerarse fundacionales y referenciales en este ámbito de reflexión se distingue claramente entre la masificación de los datos, la datificación y la digitalización de la información.

Por ejemplo, en una de las primeras obras de referencia globales a este respecto, caracterizada por Morozov (2019) como la obra *fundacional* del tema y la más citada en el mundo académico, como pueda ser «Big data. La revolución de los datos masivos», de Viktor Mayer-Schönberger y Kenneth Cukier (2013), una gran parte de la reflexión desarrollada se basa en la mencionada distinción.

Para Mayer-Schönberger y Cukier, la *datificación* es muy diferente de la *digitalización*. Mientras la primera se relacionaría con la producción de instrumentos de medida en general, la segunda sería el proceso históricamente más reciente y más particular asociado al desarrollo de los ordenadores que transforma la «información analógica en los ceros y unos del código binario». Proceso de digitalización que «impulsa la datificación, pero que no la sustituye» (Mayer-Schönberger y Cukier, 2013: 100-106).

Ahora bien, por más que histórica y conceptualmente se pueda distinguir entre los *datos masivos* o *macrodatos* y la *digitalización* de los mismos no deja de ser cierto que en la práctica y con el desarrollo de la creciente capacidad de la computación, de los ordenadores, la digitalización como medio e instrumento de la *datificación* y de otros procesos sociales parece que se acaba transformando en el objetivo y en el nuevo entorno de nuestras vidas de forma que lo digital, la digitalización, el Digital First, como caracterizaba un alto ejecutivo de los medios de comunicación españoles, se sitúa en el corazón de los datos masivos, del Big data... desplazando de la vista y de la reflexión el componente todavía muy importante de los datos masivos no digitales y todavía no digitalizados en más de un caso, como puedan ser los registros administrativos que han constituido una de las fuentes y de los componentes tradicionales más importantes de las estadísticas públicas (Desrosières, 2005), y olvidando la diferente naturaleza de los datos procedentes de unas y otras fuentes de información (Domingo i Valls, Bueno y Treviño Maruri, 2021).

No deja de ser una consecuencia lógica de lo anterior que se tienda, de forma creciente, a utilizar la denominación de Big data no solo como el equivalente a los datos masivos o macrodatos sino como un término que engloba los propios procesos y tratamientos algorítmicos de datos que, en algunos casos, como el llamado *machine-learning* forman parte más específicamente de los desarrollos más recientes de la Inteligencia Artificial (IA) pero que han podido dar un gran salto en su desarrollo precisamente gracias al Big data (Bastin y Tubaro, 2018). Por ejemplo, Vayre define al Big data como un «sistema socio-técnico», como una «evolución sociotécnica» que promueve el desarrollo de «sistemas de producción y de cálculo de los datos digitales (...), de almacenamiento y de tratamiento de datos y de la inteligencia artificial» (Vayre, 2014).

La misma noción de Inteligencia Artificial también parece constituir una categoría de contornos difusos que engloba tecnologías y prácticas muy diferentes, los algoritmos, el data mining, el deep learning, los motores de búsqueda y de recomendación, etc. (Chateauraynaud, 2019). De hecho, el proyecto de la IA fue lanzado en 1956 en la conferencia de Darmouth y, tras unos años de cierto olvido, fue recuperado a partir del 2015 al punto de configurarse en la actualidad como una expresión más de moda que el propio Big data (Rouvroy, 2016; Benbouzid y Cardon, 2018; Cardon, Cointet y Mazières, 2018).

Por ejemplo, en una encuesta realizada en España en julio de 2021 con una muestra de 1.400 encuestados de la que se daba noticia en eldiario.es (15/12/2021), los términos más frecuentes con los que los encuestados asociaban el término Inteligencia Artificial eran términos muy estrechamente ligados con la tecnología como «robots», «ordenadores» o «algoritmos», otros de connotación más positiva como «futuro», «progreso» o «avance» o algunos más negativos como «pérdidas de empleo», «miedo», «peligro» pasando por algunos otros relacionados con la ciencia ficción. Este conjunto de términos configuraba la siguiente nube de palabras en la que el tamaño de cada una de ellas representa su peso, su importancia relativa en el conjunto de asociaciones realizadas con el término Inteligencia Artificial.

