

BIG DATA Y DESCRIPCIÓN DENSA

Elisenda Ardévol

Universidad Abierta de Cataluña / Universitat Oberta de Catalunya
eardevol@gmail.com



RESUMEN

En este artículo me propongo considerar la relación entre la etnografía y los métodos digitales que se aglutinan bajo la denominación de Big Data, analítica de datos o minería de datos. En primer lugar, me interesa abordar qué entendemos por Big Data y que promesas y desafíos presenta para las ciencias sociales, para luego tratar cómo su relación con la etnografía se establece a partir de la distinción ya clásica en ciencias sociales entre métodos cualitativos y cuantitativos, así como las distintas metáforas que se han utilizado para marcar las diferencias entre la minería de grandes bases de datos y el trabajo de campo etnográfico (big data frente a small data o datos de superficie frente a datos profundos o densos). Posteriormente, a partir de mi propia experiencia en un proyecto donde trabajamos con estas dos aproximaciones metodológicas en el caso de los selfies en las redes sociales abordaré los problemas que surgen sobre la práctica, para finalmente tratar cuestiones alrededor de qué tipo de datos construimos, qué aspectos éticos se plantean y cómo se relacionan los patrones de datos con los patrones culturales en la analítica cultural Big Data.

Palabras clave: etnografía; big data; análisis cultural; selfie; redes sociales

ABSTRACT

In this article I propose to consider the relationship between ethnography and digital methods that are grouped under the name of Big Data, data analytics or data mining. First, I want to address what we mean by Big Data and promises and challenges presented for the social sciences, and then address how their relationship with ethnography is established from the classic distinction in social sciences between qualitative and quantitative methods, and the various metaphors that have been used to mark the differences between mining large databases and ethnographic fieldwork (big data against small data or surface data against deep or dense data). Further on, drawing on my own experience in a project where we work with these two approaches the case of selfie in Social Media, I will address the problems that arise on practice, to finally deal with questions about what kind of data we build, what ethical issues we face and how patterns of data relate to cultural patterns in Big data cultural analysis.

Keywords: Ethnography; Big Data; Cultural Analysis; Selfie; Social Media




BIG DATA: ¿UN DESAFÍO PARA LAS CIENCIAS SOCIALES?

La emergencia de las redes sociales hace apenas unos diez años ha revolucionado el estudio de los procesos y dinámicas sociales y culturales. Hoy día es posible seguir –incluso en tiempo real– las opiniones, ideas, sueños y emociones de cientos de millones de personas prácticamente en todo el mundo; podemos acceder a millones de imágenes, vídeos y textos que se comparten en la red, seguir conversaciones en las que están enzarzados millones de usuarios, conocer la música que escuchan y comparten, sus gustos de lectura, sus amigos, o dónde están en cada momento –y estos son sólo un ejemplo de la gran cantidad de datos digitales que generamos–.

No cabe duda que las tecnologías digitales actualmente presentes en la mayor parte de nuestras prácticas cotidianas a escala mundial abren un campo de posibilidades para la investigación social. Además, la trazabilidad de las interacciones de los distintos actores sociales (personas, grupos, corporaciones, bots, etc.), al producir, compartir, interactuar y organizar la información digital facilita la generación de nuevos métodos para extraer y analizar estas grandes cantidades de información. La automatización del registro y análisis de datos abre el camino a la extracción de información masiva, detallada y precisa de patrones de comportamiento individuales y colectivos que puede iluminar de una forma sorprendente nuestro conocimiento sobre las sociedades humanas, preguntándonos si estamos ante una revolución en los modos de conocer y si esto transforma la manera de entender las relaciones entre los seres humanos y con su entorno. El auge de los medios sociales, junto con los avances en las herramientas computacionales que pueden procesar cantidades masivas de datos (minería de datos o analítica Big Data) parece hacer posible un enfoque totalmente nuevo para el estudio de los seres humanos y la sociedad hasta tal punto que muchos hablan de una “revolución” en las ciencias sociales y de computación social (*social computing*) o analítica cultural (*cultural analytics*) para referirse al diseño y aplicación de software para el estudio de la actividad social y cultural a través de medios digitales. Como afirma Manovich (2011) la posibilidad de capturar los detalles de la comunicación diaria de la gente a través de Internet y de los dispositivos móviles puede llegar a crear la nueva sociología del siglo XXI. Para danah boyd y Kate Crawford (2012), efectivamente, el uso de procedimientos Big Data en las ciencias sociales configura una nueva forma de conocer que modifica nuestro objeto de estudio y cómo entendemos la sociedad y la cultura, además de suponer un giro radical sobre cómo entendemos la investigación social, modificando nuestros métodos y marcos de relación con la información y con las formas en que categorizamos la realidad.


Ahora bien, aunque epistemología y metodología van siempre de la mano, iguales procedimientos o prácticas aparentemente similares pueden albergar distintos marcos



teóricos y epistemológicos, así como distintos propósitos prácticos de conocimiento o de intervención en el mundo. Que efectivamente nuestra teoría y metodología se transformen al incorporar métodos digitales y la analítica Big Data no supone que se transformen en una sola dirección o de forma homogénea. Como veremos, se abren múltiples caminos de exploración, y entonces, los procedimientos de Big Data se abren también como campo de discusión y de batallas renovadas entre distintas formas de entender el mundo y de participar en él. En este sentido, es importante remarcar que lo que entendemos por Big Data puede tomar muchas formas según sus distintas propuestas y diversos objetivos, de manera que cada base de datos estructura y codifica los datos en función del propósito de sus creadores, categorizando, codificando, cuantificando, descartando, almacenando y procesando la información de maneras específicas. Por tanto, hay que ser cautos en la valoración de las expectativas que genera el Big Data en cuanto a una mayor comprensión de nuestro comportamiento social, o en relación a cómo esta analítica de datos puede contribuir a la generación de un conocimiento para la toma de decisiones que nos afectan bajo la premisa de que conlleve una gestión más eficiente y efectiva de nuestro entorno social y natural.

Por una parte, no podemos olvidar preguntarnos por la economía política, el orden de propiedad y la gestión de estos datos. Por ejemplo, en el caso de los datos que se generan en las redes sociales, los propios usuarios desconocen cómo se maneja la información que generan y la mayor parte de los científicos sociales tienen hoy en día un acceso limitado, a no ser que estén contratados por las grandes compañías que controlan esta actividad como Google o Facebook. Por otra parte, el simple acceso a grandes cantidades de datos no garantiza su inteligibilidad y correcta interpretación. La problematización y la visión crítica de la misma noción de Big Data es algo que comparten hoy en día una buena parte de los investigadores, tanto del ámbito de la ingeniería informática como de las humanidades y las ciencias sociales, especialmente en lo referente al acceso, la propiedad, el foco y la escala de la información que se maneja, y a las preguntas de investigación a las que se busca dar respuesta.


Lo que se entiende generalmente por Big Data es pues un conjunto heterogéneo de técnicas de procesar grandes cantidades de datos digitales. El término ha ganado una creciente importancia como instrumento de mercado y recientemente se ha extendido también en el ámbito científico y académico, que solía utilizar otras nomenclaturas. Según Robert Kitchin (2013:1-2) lo que caracteriza el Big Data es la capacidad de procesar para el análisis un enorme volumen de datos (en petabytes), que estos datos además, pueden ser generados y tratados a gran velocidad –prácticamente en tiempo real–, que pueden estar estructurados de diversas formas o incluso no estar estructurados, y que pueden ser exhaustivos, es decir, que pueden llegar representar prácticamente la totalidad de



las poblaciones o sistemas que se consideren, sin que se pierda su trazabilidad con la fuente original. Además estas técnicas permiten la relación de distintas bases de datos de variada procedencia y tipología, proporcionando una gran flexibilidad, extensionalidad (añadir nuevos campos) y escalabilidad (aumentar su tamaño con rapidez). La analítica Big Data puede partir de grandes bases de datos que se generan, por ejemplo, en nuestras transacciones electrónicas a tiempo real en una ciudad como Barcelona de forma geolocalizada, permitiendo, supongamos, identificar y comparar los patrones de compra de los turistas rusos frente a los japoneses o chinos. En este caso, los datos “están ahí” sólo hace falta hacerles las preguntas adecuadas para, por ejemplo, mejorar el servicio de las compañías financieras. En otros casos, se trata de generar datos específicos relevantes para una actividad, práctica o problemática concreta, y que sus resultados puedan ser utilizados de forma individual o colectiva. Por ejemplo, a partir de una necesidad como mejorar la calidad de vida de las personas alérgicas al polen, se pueden interconectar datos geolocalizados sobre distintos tipos de polen en el aire de un territorio y personalizar esta información en una aplicación según el tipo de alergia que sufre una persona para que evite esas zonas o tome la medicación necesaria. Otro ejemplo son los “personal data” que recogen dispositivos como los *wearables* deportivos para *runners*. Estos dispositivos pueden recoger, procesar e interrelacionar distintos tipos de datos como los kilómetros recorridos, la evolución de las pulsaciones cardíacas del corredor, el desgaste calórico según sus características físicas, etc. y ofrecer resultados a medida de cada persona sobre la alimentación a seguir o ejercicio físico necesario para estar en forma para la siguiente maratón, a la vez que comparte sus marcas con otros usuarios de la aplicación.

Así pues, la característica más relevante de la analítica Big Data no es la cantidad de datos, sino la complejidad de los datos que se pueden manejar y la posibilidad de relacionar distintos tipos de datos de forma agregada y en relación con otras bases de datos. Lo que es fundamental en Big Data son estas relaciones y los patrones que se pueden derivar de establecer conexiones entre distintos tipos y bases de datos, ya sean sobre un individuo, sobre relaciones entre individuos, sobre grupos de individuos, sobre individuos y cosas, o sobre la estructura misma de la información (boyd y Crawford, 2011:2). Así pues, la potencialidad del análisis Big Data está relacionada con esta identificación de patrones y que estos patrones puedan tener un capacidad predictiva. En este sentido, la magnitud de los datos que se manejan con técnicas de Big Data y la complejidad a la hora de mostrar sus relaciones, hace necesaria la introducción de técnicas de visualización, que si bien ya existían se han potenciado y ampliado para facilitar la interpretación de estos datos y la identificación de patrones de comportamiento.

No obstante, son muchos los académicos del ámbito de las ciencias sociales y las humanidades que están reflexionando críticamente sobre el poder de decisión que damos




a este tipo de análisis Big Data (Boellstorff, Bell, Maurer et al. 2015; Bollier, 2014; Markham, 2013, Burrell, 2012). Por una parte, los “datos” son construcciones humanas y cómo se recoge y categoriza la información es fundamental para su correcta interpretación. Por otra parte, la disposición de los datos de una forma visual corresponde a una intencionalidad, y puede ocultar otras posibilidades de lectura. Por tanto, hay que ser críticos frente a una idea de automatismo, objetividad o facticidad de los datos, así como de la neutralidad de sus visualizaciones. Un peligro de incluir Big Data en el análisis social y cultural es pues el de aceptar irreflexivamente estos procedimientos, sin examinar y reformular, si es necesario, nuestra perspectiva teórica y metodológica para su abordaje, y evitar en la medida de lo posible la importación acrítica de valores y supuestos mercantilistas o que depositan toda la capacidad de decisión en procesos automatizados. Incluso hay autores que denuncian la analítica Big Data como una ideología neoliberal o fruto del “solucionismo de Silicon Valley” por el cual cualquier problema social pasaría por una solución tecnológica (Kirschenbaum, 2014). Así pues, la incorporación de los métodos Big Data en las ciencias sociales “digitales” va a la par del desarrollo de distintas formas de incorporar estos métodos en la práctica de cada investigación concreta, de distintos modelos de entendimiento de este ámbito de investigación y de sus visiones críticas.

BIG DATA Y EL DEBATE SOBRE MÉTODOS CUANTITATIVOS Y CUALITATIVOS


Hay un error que persiste al examinar la relación de las ciencias sociales y las humanidades con la analítica Big Data, y este es la clasificación de la perspectiva social y cultural en el saco de las metodologías cualitativas (salvo excepciones como la economía y buena parte de la sociología), y de la analítica Big Data en el saco de las metodologías cuantitativas (Felt, 2016). Como afirman boyd y Crawford (2012) abunda la creencia que la investigación cualitativa está comprometida en interpretar historias –desarrolla una mirada subjetiva, mientras que la cuantitativa está en el negocio de producir hechos –supone una perspectiva objetiva. Es la naturaleza distinta de los datos (narraciones o hechos) lo que determina en última instancia el carácter de una u otra metodología y el análisis Big Data se situaría claramente del lado cuantitativo.

Sin embargo, quiero argumentar aquí que la relación entre la analítica Big Data y las ciencias sociales y humanísticas no debe reproducir los debates sobre métodos cualitativos y cuantitativos y sus relaciones de complementariedad o antagonismo. Todos los intentos de reproducir estas diferencias para defender la necesidad de la mirada cualitativa en los métodos asociados a el Big Data o para reclamar las bondades de lo cualitativo frente a la cuantificación yerran el tiro y no hacen más que perpetuar una división molesta para



ambas partes (Curran, 2013). Y esto es especialmente grave para la etnografía, pues al acomodarse generalmente dentro de los métodos cualitativos pierde su especificidad como aproximación epistemológica, reduciéndola al tipo de técnicas y datos que recoge e interpreta. Pero todos sabemos (o deberíamos saber) que los etnógrafos no sólo escuchan y cuentan historias, sino que también cuentan las unidades domésticas y las cabezas de ganado (Evans-Pritchard, 1940). Entonces se suele decir que la etnografía utiliza también a veces datos cuantitativos y que combinándolos procede con una metodología “mixta” (Nova, 2013). Se suele establecer entonces una especie de jerarquía: la aproximación cuantitativa también utiliza subsidiariamente datos cualitativos –como las entrevistas– de forma complementaria o para enriquecer sus hallazgos, de la misma manera que el enfoque cualitativo puede echar mano de la estadística, de manera una se pone al servicio de la otra. El problema entonces suele surgir a nivel epistemológico, dónde se dice, que ambos métodos chocan estrepitosamente: ¿Cómo puede el enfoque positivista del cuantitativista aceptar la perspectiva hermenéutica del cualitativismo o vice versa? Esta complementariedad o cualquier tipo de combinación suele implicar una correlación entre epistemología y método, de manera que si se trata con datos cuantitativos, la mirada debe ser necesariamente positivista, mientras si tratamos con datos cualitativos, la mirada será fenomenológica o hermenéutica. Sin embargo, la minería de datos y la analítica Big Data pueden sobrepasar estas limitaciones dicotómicas y, aunque trata con números, puede establecer relaciones más igualitarias con las metodologías no positivistas. Es decir, la analítica de datos está generando divergencias parecidas a las que encontramos en el debate cualitativo versus cuantitativo, pero en la práctica y en la teoría, las cosas no están tan claras. Encontramos, por ejemplo, un sector de la llamada “ciencia de datos” o “orientada a datos” (*data-driven*) que está llegando a postulados epistemológicos asociados a las metodologías cualitativas y a la etnografía. Así, según Rob Kitchin (2014), la ciencia de datos introduce la abducción como método de inferencia lógico que estaba ausente en el paradigma científico moderno (llamémosle positivista) basado en la deducción, y que esto supone una modificación del método científico. Sin embargo, este razonamiento olvida que la abducción lógica fue desarrollada y utilizada en el campo de la biología, la psicología y la antropología desde mucho antes (Bateson, 1972). Por lo tanto, la ciencia de datos no reconfigura una versión “tradicional del método científico proveyéndolo de un nuevo camino para la construcción de teorías” (Kitchin, 2014:6), sino que incorpora una práctica científica y una filosofía de la ciencia que ya se estaba desarrollando en muchos campos de las ciencias físicas, naturales y sociales.


La idea de una ciencia de la complejidad que lleve a modelos científicos más holísticos y sistémicos no es pues una idea nueva de la ciencia de datos, sino una propuesta que hace tiempo está en boga y que para algunos este cambio de paradigma ya se produjo en el pasado siglo (Capra, 1998). No obstante, la posibilidad que la minería o analítica de datos



–lo que se conoce como Big Data– ofrece al panorama antes descrito es precisamente la reconfiguración o disolución de la división entre métodos cualitativos y cuantitativos, ya que (al menos para ciertos autores) promete ser una vía cuantitativa para lidiar con el tipo de datos y de relaciones de una manera que en las ciencias sociales sólo los métodos cualitativos declaraban abiertamente abrazar.


Rob Kitchin (2014:9), citando a Brooks señala otra cuestión fundamental, y es como la analítica Big Data se enfrenta a lo social. Puesto que una buena parte de practicantes de este tipo de análisis siguen un empirismo ingenuo por el cual “los datos hablan por sí mismos” sin necesidad de teorías, se dice que los métodos Big Data pueden prescindir de cualquier necesidad de apoyarse en el conocimiento y discusiones teóricas de las ciencias sociales, esperando que un “nuevo” conocimiento de lo social “surja” espontáneamente de sus análisis. Esto genera una gran desconfianza en el campo de las ciencias sociales y las humanidades frente a el Big Data que, si se suma a lo que hemos dicho antes -la asimilación de “cuantitativo” igual a “positivista”, hace que desde las humanidades se rechacen las promesas de esta “nueva” ciencia de datos, argumentando en su contra la “incapacidad” de la aproximación cuantitativa para abordar la complejidad humana y lo inefable de su espíritu. Este escollo es difícil de salvar y se basa generalmente en un desconocimiento de la teoría social o en una “sociología de sentido común” por el cual lo social es el resultado de la suma de individuos y que los datos “están ahí fuera” y sólo es necesario ir y recogerlos.

En primer lugar debemos recordar que los datos no se encuentran como setas en el bosque. Los datos son un constructo teórico-empírico, aunque la teoría provenga de nuestro sentido común de que en el bosque hay setas y árboles (Casanovas, 1998). Los datos deben ser imaginados como datos antes que nada, y este proceso de imaginar los datos conlleva un componente interpretativo, de forma que cada disciplina o institución tiene sus propias normas y estándares para imaginar sus datos (Gitelman, 2013). Geoffrey Bowker lo expresa acertadamente al referirse a la creación de bases de datos en la práctica científica: “debemos abrir un discurso –dónde no exista aún- sobre las distintas temporalidades, espacialidades y materialidades que podemos representar en nuestras bases de datos, con vistas a diseñarlas para la máxima flexibilidad y apertura posible hacia una emergente polifonía y policromía. Los datos “crudos” (*raw*) son un oxímoron y una mala idea; muy al contrario, los datos deben ser cocinados cuidadosamente (Bowker, 2005: 183-184). Así pues, tanto en la ciencia de datos como en entre los practicantes de una estadística radical o en la GIS crítica vemos un reconocimiento de los datos cuantitativos no como hechos sino como constructos. Estas aproximaciones reconocen la naturaleza construida de los datos y el contexto cultural, social y político de toda investigación, acudiendo a la teoría social interpretativa o post-positivista para enmarcar sus preguntas de investigación, sus procedimientos metodológicos y sus resultados (Kitchin, 2014:9).



Resumiendo, hay que matizar la afirmación según la cual la analítica Big Data puede dar paso a un nuevo paradigma científico, ya que nos permite una nueva aproximación epistemológica para dar sentido al mundo y que esta consiste más que en probar una teoría mediante el análisis de datos relevantes, en buscar los entendimientos (*insights*) que nacen de los datos mismos (Kitchin, 2014:2). A lo que apunta este argumento es que la analítica Big Data puede funcionar como un método que, basado en datos cuantitativos, se pueda inspirar en –o coincida con– los planteamientos de un enfoque cualitativo, fenomenológico o hermenéutico. La mayor parte de los enfoques llamados cualitativos insisten en que no buscan probar teorías, sino en producir entendimientos (*insights*) a partir de la interpretación de los datos, siendo el proceso más inductivo que deductivo (Taylor y Bogdan, 1987, Denzin y Lincoln, 2000). La llamada teoría fundamentada o *Grounded Theory* (Strauss y Corbin, 1998) es un buen ejemplo de ello, así como gran parte de las formas de entender el método etnográfico hoy en día (Ferrándiz, 2011). Además, la analítica Big Data parece ofrecernos una mirada distinta a los datos cuantitativos de la que nos ha proporcionado la estadística hasta ahora (Pink, Markham, Arkhama, 2015), ya que pone el énfasis no en la representatividad de la muestra y en la búsqueda de relaciones causales, sino en la significación de la muestra y en las correlaciones entre variables para llegar a una mayor comprensión del fenómeno estudiado (Zwitter, 2014). Ciencia de datos y enfoque hermenéutico coinciden en que no se trata de demostrar teorías, sino de hallar patrones relevantes para la investigación. Este tipo de aproximaciones epistemológicas en la ciencia de datos confluye con la perspectiva cualitativa, y muy especialmente con la práctica etnográfica de muchos investigadores e investigadoras que se proponen como objetivo la búsqueda e identificación de patrones (culturales) para construir sus modelos explicativos.

Heather Ford (2014) que ha trabajado como etnógrafa con analistas de datos para la Wikipedia, nos dice que en la práctica, hay una comprensión mutua porque en ambos campos es necesario sumergirse en los datos para buscar un sentido. Otro ejemplo de la amalgama entre etnografía y analíticas Big Data puede encontrarse en la “etno-minería”, metodología desarrollada por Aipperspach et al. (2006) que supone un trabajo conjunto entre etnógrafos, ingenieros y analistas de datos en relación con determinar qué interpretaciones de los datos son las más interesantes para la pregunta de investigación, y por tanto, cómo deben caracterizarse, qué datos deben tomarse y cómo deben seleccionarse y transformarse para llegar a interpretaciones relevantes. En este estudio sobre la incorporación de las tecnologías digitales en el hogar, para determinar qué tipo de datos se construyen se parte de una observación etnográfica que dé unas primeras impresiones sobre qué es significativo en esa situación, y de forma iterativa, se desarrolla el proceso de recogida de distintos tipos de datos con sensores que captan los movimientos de las personas en la casa, programas de seguimiento de las acciones en sus ordenadores portátiles, registros visuales, observación directa y realización de entrevistas. Es la interrelación de estos datos




heterogéneos y las categorías de análisis que van emergiendo los que utilizan para validar de forma también iterativa, las interpretaciones de los investigadores. También es interesante el estudio que realizan el equipo de ingenieros y antropólogos de “Digital Living” en el cuál la etnografía sensorial orienta la investigación sobre cómo mejorar el consumo energético en el hogar a partir de seguir las rutinas cotidianas para desarrollar aplicaciones digitales a partir de observar los patrones de comportamiento de los usuarios y el sentido que dan a sus acciones, no sólo en relación al ahorro energético sino también a su entendimiento de lo que es un ambiente hogareño. Aunque el desarrollo de sensores y las métricas de consumo poco tienen que ver con la técnica de las entrevistas y la observación participante, los miembros del equipo comparten una misma epistemología, reflexiva y abierta hacia las contingencias del proceso de investigación, sensible a los matices y relaciones en el campo, produciendo interpretaciones contextualizadas que tienen sentido tanto para los investigadores como para los participantes en el estudio (Pink, Markham, Arkhama, 2015).

En definitiva, la analítica Big Data y los métodos llamados cualitativos pueden convivir creando marcos epistemológicos compartidos que definan el problema de investigación y lo abran para la exploración conjunta (Curran, 2013). Entender los datos en el Big Data como meramente cuantitativos y los datos etnográficos como exclusivamente interpretativos conduce a construir una tipología abstracta de datos que impone una diferencia ontológica que luego determina una epistemología que actúa como mecanismo identitario de una disciplina o de un método, o piénsese en sentido inverso. En todo caso, y ya centrándonos en la etnografía, esta se caracteriza precisamente por la heterogeneidad de sus datos y su gran flexibilidad e imaginación en el desarrollo de procedimientos para obtenerlos, analizarlos e interpretar los hallazgos. Aspectos altamente presentes también en la analítica Big Data.


DATOS PEQUEÑOS, DATOS DENSOS, DATOS, DATOS...

La utilización de metáforas puede ser muy útil para establecer puentes de diálogo pero a veces los puede entorpecer. Me resisto a hablar de “los” Big Data como “datos grandes” porque entonces la etnografía se caracteriza por contraposición, por trabajar con datos “pequeños” (Burrell, 2012). Tanto en la analítica Big Data como en etnografía, una sola imagen puede analizarse de formas muy complejas, como millones de datos pueden analizarse de formas muy simples. En esta dirección es útil señalar el trabajo de Wendy Hsu en el campo de la etnomusicología. Esta antropóloga incorpora en su práctica etnográfica analíticas Big Data tanto para geolocalizar millares de seguidores en la red de grupos de rock asiáticos como para realizar un análisis en profundidad de una sola grabación de sonido (Hsu, 2014).



Otra posición común es considerar que los datos que maneja la analítica Big Data son datos “descontextualizados” siendo la función de las ciencias sociales el “contextualizarlos”, ya que para entender ciertos patrones de comportamiento de los datos es necesario recurrir a información contextual de tipo social, histórico y cultural. En este sentido, se dice que “los grandes datos” pueden ampliar las posibilidades de análisis y permitir nuevas aproximaciones y técnicas, pero nunca podrán reemplazar los estudios tradicionales de los “datos pequeños” (*small data*) (Kitchin, 2014:9). Entonces, los grandes datos sólo pueden proveernos de descripciones superficiales, instantáneas interesantes, pero no pueden penetrar en las geografías de la raza, el lenguaje, explicar el porqué de las aglomeraciones y las segregaciones en ciertos lugares particulares, aspectos que sólo los “pequeños datos” basados en un análisis en profundidad y a pequeña escala nos puede proveer. Cuando Lev Manovich defiende el uso de datos masivos en las humanidades y en las ciencias sociales, alude también a la contraposición entre “datos superficiales” y “datos profundos” al comparar las limitaciones y el alcance de Big Data frente a la etnografía. Insiste, en este caso, en que la diferencia no está en que los etnógrafos obtengan unos pocos “datos profundos” (*deep data*) sobre unas pocas personas frente a los “datos superficiales” (*surface data*) sobre montones de gente que puede obtener la analítica Big Data, sino que ambos tipos de datos pueden considerarse “profundos” sólo que llegan a un tipo de conocimiento distinto. Se hacen distintas preguntas, identifican distintos patrones y llegan a distintos entendimientos (*insights*) (Manovich, 2011:17). La cuestión se dirime argumentando que en la relación cara a cara, el etnógrafo descubre ciertos patrones que nunca captará el analista Big Data en sus visualizaciones, y por el contrario, el analista descubrirá patrones que pasan desapercibidos a los ojos del etnógrafo sobre el terreno. Esta es una distinción interesante y volveremos sobre ella.

El problema de estas metáforas es que presuponen que tanto los datos “grandes” como los “pequeños” no llevan teoría, es sólo una cuestión de contexto, de detalle o de profundidad. Tricia Wang (2013) nos quiere llamar la atención frente a los peligros de tratar los datos como si no llevaran teoría. En su caso, también utiliza una metáfora para contraponer los grandes datos a los “datos densos” (*thick data*) en honor de Clifford Geertz y su aproximación interpretativa a la cultura (1991). Clifford Geertz utiliza la contraposición entre descripción plana o superficial y descripción densa de Gilbert Ryle para explicar la naturaleza del dato etnográfico. Un guiño puede ser descrito como el parpadeo intermitente de un ojo mientras se mantiene el otro abierto. Esta es una descripción de un patrón de comportamiento observable que se da en muchas sociedades humanas y puede ser una descripción adecuada, pero no suficiente para constituirse en una descripción “densa”. El “parpadear con un ojo mientras se cierra otro” es un patrón cultural si y sólo si tiene una significación culturalmente concreta, que puede variar de un marco cultural a otro y de una situación a otra. De la correcta interpretación o no de ese comportamiento como un guiño o como un



tic puede depender la vida de alguien, por ejemplo. Establecer que el significado cultural de ese parpadeo puede ser un guiño, si es de complicidad, o un tic, si se considera que el movimiento es involuntario, y qué lugar ocupa el guiño en determinado entramado cultural y situación, es lo que constituye la base para una descripción densa. La analítica Big Data puede examinar patrones de comportamiento concreto, como por ejemplo, analizar millones de fotografía en busca de un ojo cerrado y otro abierto, y comparar estas fotografías con las que no presentan ese rasgo para establecer conexiones formales de similitud o diferencia, pero sólo una descripción densa puede revelar el significado cultural, si tiene alguno, de estas conexiones. Según Tricia Wang, la analítica Big Data produce descripciones “planas”, y sólo la etnografía produce descripciones “densas”, descripciones que generalmente no se expresan de forma numérica sino en forma de narraciones, de historias. Si sólo nos centramos en los patrones de comportamiento que nos revelan los grandes números, y nos fiamos exclusivamente de ellos, podemos llegar a tomar decisiones basadas en “descripciones planas”. Y si además no contamos con los “datos densos” a la hora de crear nuestros algoritmos para optimizar nuestras métricas, no sólo el conocimiento social y cultural, sino las personas y sus experiencias van a quedar fuera de la ecuación y nos quejaremos, con razón, del poder ciego de las máquinas sobre nuestras vidas. Esta autora quiere evitar a toda costa esta alienación de lo humano y revalorar la oferta etnográfica cómo método indispensable para la inteligibilidad de los Big Data contraponiendo los números a las historias, la importancia de la narración frente a la categorización. Mientras que la categorización ordena el mundo –lo cierra, la narración lo despliega; abre mundos posibles (Wang, 2013)–.

Sin embargo, la práctica etnográfica opera con categorizaciones al igual que es creadora de historias. Para lidiar con esta cuestión John Curran (2013) nos propone girar la vista atrás, hacia la historia de la antropología y el surgimiento del método etnográfico tal y como lo entendemos ahora. Según este autor, Bronislaw Malinowski –considerado como el fundador del actual método etnográfico, no pretendía tan sólo generar narraciones etnográficas significativas, sino que buscaba con ello desarrollar un nuevo método científico que fuera robusto para generar modelos sociales y pudiera apoyar su teoría funcionalista, y esto a través de generar una metodología que le permitiera la interrelación de diversos tipos de datos (estadísticas, mapas, gráficos, tablas de parentesco, etc. pero también la observación directa sobre lo que se hace y anotaciones sobre lo que se dice), siendo lo más importante de este nuevo método que el investigador tuviera un contacto directo con la sociedad que estudiaba; la importancia del “estar ahí” como elemento fundamental para una comprensión holística de los patrones culturales de un determinado grupo social. Malinowski publica en 1922 su monografía sobre las islas Trobriand argumentando su rechazo a basar la ciencia antropológica exclusivamente en datos de “segunda mano” generados por los agentes coloniales, viajeros, mercaderes o misioneros, o basados exclusivamente en




entrevistas y censos de población; con ello no pretendía alejarse de la ciencia, sino realizar una contribución substantiva para mejorar la práctica científica en su campo. Su intención era construir modelos sociales que se fundamentaran tanto en la estadística como en el registro de observaciones e impresiones del investigador en un diario de campo, incluyendo como imprescindible el registro de las narraciones de los propios sujetos de la investigación y sus entendimientos. Si consideramos todo esto, no se trata de la obtención de datos “pequeños” o de datos “profundos” frente a datos “grandes” o “superficiales”, sino de la creación de un entramado complejo de datos heterogéneos que se interpretan a la luz de la experiencia del etnógrafo de haber palpado esa realidad. Lo que sostiene John Curran (2013), como etnógrafo que trabaja para la industria y con Big Data, es que hay que dejar de pensar en Big Data sólo como “datos grandes”, código y algoritmos, como algo que la etnografía “no es”, para pensar dónde la interpretación cultural y la analítica Big Data se encuentran (Curran, 2013:68-69).

Como nos recuerdan Sarah Pink, Deborah Lupton y otros autores (2016), la analítica Big Data no es sólo un método de investigación, no sólo estructura nuestro conocimiento de la realidad y orienta nuestras decisiones, también ordena nuestra realidad misma y esto nos afecta a todos, nos interpela emocionalmente y se enreda con nuestra experiencia de estar en el mundo. El Big Data no son sólo códigos y algoritmos, son también emociones, promesas, frustraciones, miedos, deseos –son parte de nuestra experiencia, tanto si somos analistas de datos como si somos científicos sociales, políticos, empresarios, trabajadores, médicos, pacientes, deportistas, etc. Una aproximación etnográfica a Big Data debe incluir también este aspecto experiencial no sólo en las vidas de los ciudadanos “corrientes” o de los “usuarios”, sino también en los procesos de diseño de las bases de datos, en el desarrollo de los algoritmos, en la forma de crear código, generar, almacenar, tratar y cuidar de los datos (Pink, Markham, Akhama et al., 2016).

SELFIES, NARRATIVAS PERSONALES Y BASES DE DATOS

Cuando planteamos en nuestro equipo de investigación el proyecto *Selfie stories y personal data: metodologías híbridas para el análisis de las narrativas visuales en la cultura digital** nuestra propuesta era precisamente explorar las posibilidades de la analítica de datos en relación con la etnografía proponiendo una aproximación “mixta” de métodos cuantitativos y cualitativos para el análisis cultural de los “selfies” en las redes sociales. Decidimos centrarnos en el “selfie” en las redes sociales como Twitter e Instagram como un exponente actual de los modos de presentación de uno mismo que ha cobrado una especial popularidad, convirtiéndose en un fenómeno social a escala global. Considerábamos que muy a menudo,




los medios de comunicación han relacionado esta práctica con el narcisismo y exhibicionismo, resaltando los aspectos más negativos o sus consecuencias peligrosas (como caerse de un acantilado por hacerse un selfie), protagonizando innumerables polémicas en este sentido, en las que se ven implicados tanto personajes públicos como *celebrities* y las personas corrientes (San Cornelio, 2015). Mientras que, por otra parte, diversos estudios recientes muestran que los selfies forman parte de las narrativas transmedia contemporáneas (Fallon, 2014) o que deben comprenderse como parte de una conversación digital (Gómez y Thornham, 2015; Gunthert, 2015).

Focalizar nuestro estudio en el “fenómeno selfie” nos permitía indagar sobre la producción de “datos personales” susceptibles de tratamiento analítico Biga Data, a la vez que explorar su componente narrativo vinculado a la construcción de la identidad personal y las problemáticas que suscitan temas como el acceso a los propios datos, la intimidad y la privacidad en el contexto de las tecnologías digitales. El objetivo inicial era examinar esta práctica cultural combinando el enfoque cualitativo y etnográfico con una analítica Big Data para experimentar su alcance y limitaciones, a la vez que reflexionar sobre las distintas lógicas narrativas y de archivo que entrecruzan en esta manifestación digital.


Para explorar la lógica de las grandes bases de datos en relación con el análisis cultural podemos empezar de una manera sencilla, por ejemplo, tecleando la palabra “selfie” en el buscador de imágenes de Google, obtenemos un conjunto de millares imágenes cuyo nombre de fichero contiene este término. Si realizamos esta misma búsqueda en Twitter o Instagram, nos mostrará las imágenes que los usuarios han etiquetado así con un hashtag o los usuarios cuyo nombre contiene este término. Las bases de datos online nos proporciona diferentes formas de recuperar grandes cantidades de información a partir de las distintas clasificaciones y categorizaciones que estructuran cada base de datos. Según Manovich, el buscador produce algo similar a una narrativa no lineal en el sentido de que muestra múltiples imágenes que han sido categorizadas de la misma manera, y por tanto, bajo el mismo concepto o idea (Manovich, 2007). Pero, ¿configura este conglomerado alguna estructura con sentido? Instintivamente, seleccionaremos algunas de ellas, que siguen patrones comunes, y descartaremos otras que se alejan demasiado o producen “ruido”, estaremos “limpiando” nuestra muestra para adaptarla a este concepto previo del “selfie” como un auto-retrato realizado con una cámara digital –generalmente incrustada en un dispositivo como un smartphone o tablet.

Otra forma de proceder sería que pudiéramos explorar la red a partir del reconocimiento automático de imágenes. Que el buscador sólo seleccionara aquellas imágenes que cumplen ciertos requisitos formales, por ejemplo, que incluyan un rostro y que además tengan un ángulo y enfoque determinado para ser técnicamente un “selfie”. En este caso




estamos buscando un patrón visual independientemente de cómo haya sido etiquetado por las personas que lo han subido a la red. Esta idea se aproxima a la propuesta de Manovich y su equipo en el proyecto *Selficity* (2014). Este autor entiende la “analítica cultural” (*cultural analytics*) como la exploración y la investigación de conjuntos de datos culturales masivos de material visual -tanto artefactos visuales digitalizados como medios visuales e interactivos contemporáneos. Con el fin de asumir el reto de cómo explorar mejor las grandes colecciones de datos de un rico contenido cultural, su equipo de investigación ha desarrollado herramientas extractoras y visualizaciones de alta resolución para el procesamiento de imágenes digitales. El proyecto al que me refiero aquí construye una base de datos cerrada que incluye 120,000 fotografías publicadas en Instagram geolocalizadas en cuatro ciudades: Berlín, Moscú, Sao Paulo y Nueva York. Las imágenes fueron seleccionadas por cumplir las categorías formales de un selfie (que aparezca un rostro, cierta inclinación del cuerpo por sostener una cámara, etc.) y categorizadas manualmente según una tipología predefinida (localización, sexo, edad, posición del rostro, expresión facial, etc.). Las visualizaciones son interactivas, de forma que podemos escoger, por ejemplo, cierta característica (hombres con gafas y sonriendo) y ver su peso estadístico dentro de la muestra y como se distribuye esta imagen por ciudades. Así se pueden extraer ciertas conclusiones, como por ejemplo, que hay más selfies de mujeres que de hombres, o que en Sao Paulo la gente se muestra mucho más sonriente que en Moscú, o que apenas hay selfies con gafas en Berlín. Podemos preguntarnos ¿Qué pasa en Sao Paulo que la gente parece más contenta? ¿Por qué la gente no se hace selfies con gafas en Berlín? La base de datos de *Selficity* nos permite explorar los límites y las posibilidades de la analítica Big Data para las ciencias sociales en un entorno controlado, y en este sentido, es muy sugerente para introducirse de forma experiencial en qué nos puede ofrecer una analítica de este tipo.

Algunas cosas que podemos aprender. Primero, que la selección de fotografías responde a una categorización previa del analista y es necesario una limpieza de los datos “en bruto” captados por la herramienta extractora (API). En este caso, se parte de un concepto o idea previo de “selfie” independientemente de cómo se hayan etiquetado las imágenes. Segundo, que la categorización de las imágenes (manual o no) es también una decisión analítica. En este ejemplo, como señala Losh (2014), la base de datos excluye otros géneros que no sean el masculino y el femenino, es decir sólo contempla la dicotomía sexual macho o hembra y no se tiene en cuenta otras posibles autoclasificaciones de las personas como por ejemplo, gay, queer, bisexual, transexual, pansexual, etc. Tercero, en este caso, las imágenes se analizan en función de las categorías analíticas consideradas como “hechos” objetivos y a partir de estos se infiere un conocimiento cultural, como que en Sao Paulo las personas son más felices que en Moscú, porque hay un mayor número de personas sonrientes



en la muestra. Este tipo de inferencias no son autoevidentes ni constituyen un análisis cultural sin riesgos. Cuarto, que para la selección de estas imágenes se ha estimado que al estar abiertas y disponibles (no protegidas por los usuarios con la opción de privacidad que las aplicaciones ofrecen) son públicas, y por tanto, no se han considerado necesario contemplar cuestiones de privacidad o protección de la intimidad. Finalmente, *Selfiecity* nos plantea otra reflexión: Plantea que se trata de una analítica cultural sólo porque los datos (las imágenes) que se utilizan son categorizados como “culturales” (son un producto simbólico como podrían ser las obras de arte, música, películas, poesía, etc.), no porque se aplique algún tipo de “análisis cultural”. De una categorización puramente formal del selfie no puede derivarse un comportamiento cultural no ambiguo y homogéneo.


Cuándo nos enfrentamos a cómo construir nuestro modelo analítico Big Data en torno a los selfies, la primera duda que surgió fue precisamente cómo definíamos este universo: ¿a partir de la estructura formal de la imagen como “selfie”? ¿A partir de las taxonomías de los propios usuarios que etiquetan la imagen como “selfie”? Como hemos señalado antes, una primera aproximación a las redes sociales focalizada en el uso del selfie nos revela que la gente puede etiquetar como “selfie” cualquier cosa (incluyendo imágenes porno, gatitos, flores, pasteles, zapatos o auriculares), y por otra, no siempre que la gente se hace un selfie lo etiqueta como tal, de manera que hay muchos selfies en la red que no están categorizados como selfies por los propios usuarios y usuarias. Además, las personas somos creativas y desarrollamos infinitas variaciones de selfie/selfies en distintos idiomas que señalan hacia distintas intencionalidades o modalidades de selfie como por ejemplo selfieporquesi (sin ningún motivo), selfiegram (selfie de Instagram), belfies (selfies de culo), o incluso nuevos términos que derivan del “selfie” como birthies (selfies de cumpleaños), etc. También encontramos categorías de selfies que apuntan a un componente de género (selfiegay, selfiequeer) o que responden a una campaña activista como por ejemplo, #blackselfieday en los Estados Unidos, el #selfieforpeace realizada en Pakistan y la India, o la campaña contra el cáncer #nomakeupselfie originada en el Reino Unido. El formato selfie también se utiliza como protesta política aunque el término mismo no esté en la etiqueta, como la campaña #meninjihad donde muchos hombres se auto-fotografían con el jihat para protestar por el uso obligatorio de esta prenda para las mujeres en determinadas localizaciones y grupos religiosos, o el movimiento #yosoy132 como protesta estudiantil en México. Así pues una primera aproximación al universo selfie ya nos da una idea de la importancia del contexto y de la intencionalidad o la motivación para hacerse un selfie y de la gran creatividad que se despliega en torno a esta modalidad del auto-retrato. Lo que salta a primera vista es la importancia de las narrativas que acompañan el selfie, la necesidad de contemplar para un análisis cultural la nube de etiquetas en la que se inserta, los comentarios que se hacen y los que reciben.



Consideramos que una analítica Big Data que quiera contemplar un análisis cultural de los selfies debería tener en cuenta las dinámicas sociales y culturales en las que se inserta y las propias autodescripciones de los usuarios a la hora de desarrollar las categorías descriptivas, la herramienta de extraer datos e interpretar sus resultados. En nuestra investigación decidimos pues no centrarnos exclusivamente en el propio reconocimiento de patrones relativos al componente visual, como en el caso del estudio antes citado, sino analizar tanto en el análisis Big Data como en el trabajo de campo etnográfico, cómo se inscriben los selfies en las narrativas personales y colectivas, teniendo en cuenta cuándo las personas deciden hacerse un selfie, cuándo deciden etiquetarlo como tal, inventarse etiquetas derivadas, o sencillamente no identificarlos como un selfie y utilizar otras etiquetas relativas a otros aspectos que consideran que define mejor la fotografía como el acontecimiento, la localización o el momento en el que se toma la fotografía, etc. Para ello adoptamos una doble estrategia: desarrollar un extractor de datos que tuviera en cuenta las etiquetas (*hashtags*) y comentarios, y que también pudiera recoger la línea temporal personal o colectiva, en torno a la foto, un acontecimiento o un tema. También entendíamos que para realizar un análisis cultural desde una aproximación etnográfica debíamos usar distintas estrategias de aproximación apoyando desde el inicio la analítica Big Data en la observación del comportamiento de los usuarios fuera y dentro de la red (Blok y Pedersen, 2014), atendiendo tanto a las prácticas cotidianas dónde se toman las imágenes como a las tramas de significación en las que el selfie se inserta una vez se sube a la red. En esta dirección, suponíamos que la categorización de la imagen por parte de los usuarios era una práctica cultural consustancial y una información fundamental para la comprensión del fenómeno selfie. Y que era fundamental tener en cuenta también el contenido y aspecto formal de la fotografía, ya que es en el juego (transgresivo o no) con las convenciones del selfie dónde el comportamiento adquiere un sentido cultural vivo.


Para probar nuestro modelo elegimos un evento musical multitudinario como el Primavera Sound en Barcelona. La ventaja de escoger este acontecimiento cercano es que teníamos acceso al evento, podíamos conversar con la gente y entrevistarla, así como seguirla online y, en definitiva, podíamos realizar una observación participante directa, desplazándonos al sitio del evento, al mismo tiempo que el extractor recogía datos de la actividad online etiquetada con el hashtag del mismo. De este modo podíamos interpretar los datos obtenidos por la herramienta a la luz de nuestra experiencia directa en el campo y situar las preguntas que podemos hacer a las visualizaciones y a los patrones que revelan.

Para crear las visualizaciones fue fundamental el trabajo de campo realizado. Si bien es cierto que el analista puede jugar a cruzar los datos de todas las formas aleatorias posibles, lo cierto es que se tiende a buscar cierto sentido en los cruces entre variables, entre otras cosas, para ahorrar tiempo y esfuerzo. Así, no es que las visualizaciones de los datos “confirmen” nuestras impresiones de campo, sino que nuestras impresiones de campo



guían la elaboración de las visualizaciones y contribuyen a confirmar o no si hay patrones relevantes tal y como sugerían nuestras impresiones en el campo. Por ejemplo, observamos un patrón de comportamiento en este festival era hacerse un selfie individual o colectivo delante de la entrada, mientras que son mucho más escasos los selfies producidos durante el concierto, cuando se realizan muchas más fotos, pero del escenario. El comportamiento de los datos en el tiempo corroboraba esta impresión. Sin embargo, el análisis de los datos producidos online (retuits) “revela” aspectos difíciles de ser observados directamente, como la interrelación entre moda y música, destacado por el mayor número de retuits que generan las marcas y las youtubers o blogueras de moda presentes en el evento. Para interpretar el patrón de comportamiento de “hacerse un selfie” teníamos también las entrevistas realizadas a los participantes, así como los hashtags y comentarios subidos a la red junto con el selfie. Además, para incorporar la experiencia personal en nuestra base de datos lo que hicimos fue hacernos fotos también nosotros y subirlas a la red con el mismo hashtag del evento. Mediante este movimiento reflexivo de incluirnos a nosotros mismos en la base de datos lo que pretendíamos era evitar en la medida posible “datificar” a la gente que había participado en el evento, estableciendo así una suerte de compromiso deontológico.


Experimentando con nuestros primeros resultados en el caso del festival de música, podemos apreciar como no hay una diferencia entre cómo se enfoca el trabajo de campo en el evento y cómo se tratan y se interpretan los datos en las visualizaciones Big Data, sino que los distintos procedimientos siguen una misma lógica de investigación. La ventaja de estas visualizaciones desde una perspectiva etnográfica, y comparándolo con nuestra experiencia en el uso de otros procedimientos de tipo estadístico, es la posibilidad en todo momento de ir al caso particular, es decir, no perder de vista la singularidad; en nuestro caso, la posibilidad de recuperar en cada momento de la visualización de los patrones generales, la imagen concreta, el autor, las etiquetas y sus comentarios. En este sentido compartimos la opinión de Wendy Hsu (2012), para la cual los datos digitales sirven para confirmar o no los patrones culturales observados en el campo y son útiles en cuanto se materializan en una visualización para explicitarlos y facilitar el juego interpretativo. Para esta autora, la posibilidad de visualizar los datos a diferentes escalas y poder ver el comportamiento particular tanto como el general supone una forma nueva de conectarse sensorialmente con la materialidad del trabajo de campo (Hsu, 2012). Estas formas de visualización computarizada, nos dice, son interesantes porque nos ofrecen los datos de campo de una forma recontextualizada, enfatizando su materialidad. Si bien es cierto que hay programas de tratamiento textual que nos ofrecen procesos automatizados de codificación y categorización que nos son muy útiles, por ejemplo, para el análisis de contenido de las entrevistas, la analítica Big Data nos ofrece aproximaciones no textuales de gran interés etnográfico. Algo parecido nos sucedió a nosotros cuando observamos las distintas visualizaciones de los datos del Primavera Sound. Por ejemplo, en el gráfico



temporal, cada foto era representada por un punto y coloreada según su tipología, de manera que siguiendo la línea temporal veíamos a qué momento del concierto pertenecía, y por tanto, nos trasladábamos de alguna manera allí, materializando nuestras impresiones en el campo. Nuestras notas de campo, nuestra memoria, y nuestras visualizaciones de los datos en la pantalla estaban trabajando conjuntamente para la interpretación de la experiencia. Situar la analítica Big Data como parte constitutiva de la práctica etnográfica es clave para la etnografía digital (que incorpora métodos digitales) porque conecta los distintos tipos de datos (numéricos, textuales, visuales, sonoros, geográficos, etc.) en un mismo marco interpretativo abriendo el espacio analítico para la producción de sentido; para la producción de una descripción densa o análisis cultural.

PATRONES DE DATOS Y ANÁLISIS CULTURAL

La idea de que la analítica Big Data necesita a las ciencias sociales para desvelar los porqués del comportamiento humano sólo es cierta si la analítica Big Data se propone desvelar los porqués del comportamiento humano. Como bien decía Manovich (2011), el tipo de conocimiento que buscan es distinto. El objeto de estudio de la ciencia de datos no son las personas, ni el comportamiento humano, sino el comportamiento y naturaleza de los datos. Por tanto, y por definición, no puede descubrir nada sobre lo humano, lo social o lo cultural, sólo puede saber sobre datos y comportamiento de datos. Que los datos los genere una persona, un estornino, un mosquito o un sensor es algo que no creo que les deba preocupar demasiado. Se puede tomar el vuelo de los estorninos y modelizarlo, e incluso predecirlo, aun sin saber por qué ni para qué vuelan de la manera en que lo hacen. Decir que lo que porta la ornitología a los Big Data es el porqué no tiene mucho sentido. Lo que puede aportar la ornitología a los Big Data no son porqués sino preguntas y quizás esas preguntas harán avanzar a la ciencia de datos hacia nuevas formas de codificar, categorizar y entender los datos. Lo mismo creo que sirve para el caso de las ciencias sociales. Modelizar el comportamiento humano no significa hallar sus leyes. Así como la ciencia de datos busca patrones de datos, las ciencias sociales buscan patrones sociales y culturales. La interpretación del comportamiento de los datos como si fuera el comportamiento real de los individuos y los patrones de datos que hallamos como patrones culturales de una sociedad no es autoevidente ni está justificado. Una cosa es identificar patrones y otra es que esos patrones de datos puedan traducirse a patrones culturales, es decir, no se trata de que el análisis cultural “explique” los patrones identificados mediante el análisis Big Data, sino que éstos tengan sentido desde un análisis cultural. A ningún analista Big Data se le ocurriría buscar patrones significativos en millares de distintas poblaciones de bacterias sin tener al lado un biólogo que guiara sus codificaciones e interpretara sus hallazgos, pero



buscar patrones formales en los “datos culturales” como millones de fotografías de internet y decir que se ha encontrado un patrón cultural que va a revolucionar cómo entendemos las emociones humanas, eso aparece como sorprendentemente factible.

Podríamos concluir con John Curran (2013:71) que una etnografía digital que incorpore analítica Big Data es posible y que esto supone abrir un campo de investigación donde la etnografía y la analítica Big Data no establezcan distinciones a priori por el tipo de datos que utilizan o por la naturaleza intrínseca de su aproximación epistemológica. Los datos etnográficos no se caracterizan por su soporte físico o digital, ni por su naturaleza cualitativa o cuantitativa. ¿Dónde está el dato etnográfico? Un dato etnográfico puede ser cualquier cosa que la etnógrafa decida que es un dato etnográfico. Es decir, cualquier diferencia que marque una diferencia que le ayude a distinguir patrones culturales y hacer inteligible ese entorno social que busca comprender. El dato etnográfico es un logro, el producto de una conexión entre distintos tipos de datos, la memoria de la experiencia del etnógrafo en el campo y sus categorizaciones y marcos teóricos. Es fundamentalmente la introducción de la experiencia, el “estar ahí” lo que distingue al dato etnográfico de otros tipos de datos, y es lo que convierte un registro audiovisual, una anotación, una serie temporal de fotografías o una nube de etiquetas en un dato etnográfico capaz de contribuir a una descripción densa. La introducción de analíticas Big Data genera puede generar un nuevo tipo de datos etnográficos que no son accesibles mediante la observación directa, pero que se relacionan con la observación directa y con la propuesta teórica para construir su significación y su relevancia para el estudio.

Una última reflexión. La analítica Big Data no es inocua para la etnografía, plantea nuevos retos sobre cómo se relaciona el etnógrafo con el campo y con las personas que participan en su investigación, y tiene repercusiones en las asunciones que hacemos sobre la responsabilidad individual y las relaciones de poder (Zwitter, 2014). En el Big Data, las personas que participan en la investigación son millares sino millones, y podemos acceder de forma indirecta a la expresión de sus gustos, creencias, afinidades, opiniones políticas, sentimientos. Esto no sólo plantea un problema ético relacionado con la protección de datos, la privacidad o la intimidad, o en cómo obtener el consentimiento informado o anonimizar los datos, sino de ética entendida como el tipo de relación que establecemos con nuestros participantes cuando el campo incluye también grandes bases de datos. El trabajo de campo etnográfico supone una relación con las personas y un compromiso con ellas. No estudiamos a “sujetos”, sino que investigamos “junto” a personas, de manera que lo que tengamos que decir, no lo decimos “sobre” ellas, sino sobre nuestro problema de investigación y lo que ellas nos aportan para resolverlo (Ardévol, 1998). ¿Cómo podemos hacer una “antropología compartida” (Rouch, 1975) con/de Big Data? Paradójicamente, sólo experimentando con Big Data podremos afrontar etnográficamente estas cuestiones.


NOTA

* El proyecto *Selfie stories y personal data: metodologías híbridas para el análisis de las narrativas visuales en la cultura digital* (<https://selfiestories.net>) es un proyecto de investigación del grupo de investigación en cultura y comunicación digital *mediacions* de la Universitat Oberta de Catalunya financiado por la Fundación BBVA en la I convocatoria de proyectos de investigación de la temática Humanidades Digitales en 2014. Participan Gemma San Cornelio, Antoni Roig, Elisenda Ardévol, Begonya Enguix, Pau Alsina,

Irma Vilà Òdena, Edgar Gómez Cruz, Matteo Ciastellardi, Marco Quaggiotto, Erick Gonzalo Gómez Narváez, y Teresa García Català.

REFERENCIAS

- Aipperspach, R. (2006). *Ethno-Mining : Integrating Numbers and Words f rom the Ground Up*. Retrieved from <http://www.eecs.berkeley.edu/Pubs/TechRpts/2006/EECS-2006-125.html>
- Ardévol, E. (1998). Por una antropología de la mirada. en Luis Calvo. *Perspectivas de la Antropología Visual*. Revista de dialectología y tradiciones populares del CSIC. Madrid.
- Bateson, G. (1972). *Steps to an ecology of mind: Collected essays in anthropology, psychiatry, evolution, and epistemology*. University of Chicago Press.
- Baym, N. K.. Data not seen: The uses and shortcomings of social media metrics. *First Monday*, [S.I.], sep. 2013. ISSN 13960466. Available at: <<http://firstmonday.org/ojs/index.php/fm/article/view/4873/3752>>. Date accessed: 13 Oct. 2016. doi:10.5210/fm.v18i10.4873.
- Blok, A., & Pedersen, M. A. (2014). Complementary social science ? Quali-quantitative experiments in a Big Data world, (December), 1–6.
- Boellstorff, T., G. Bell, B. Maurer, M. Gregg and N. Seaver. (2015) *Data, Now Bigger and Better!* Prickly Paradigm Press.

- 
- Bollier, D. (2010). *The promise and peril of Big Data*. The Aspen Institute: Washigton.
- Bowker, G. C. (2005) *Memory Practices in the Sciences*. MIT Press, Cambridge, Massachusetts.
- Boyd, D., & Crawford, K. (2012). Critical questions for Big Data. *Information, Communication & Society*, 15:5(February), 662–679.
- Burrell, J. (2012). The Ethnographer's Complete Guide to Big Data: Small Data People in a Big Data World (part 1 of 3). <http://ethnographymatters.net/blog/2012/05/28/small-data-people-in-a-big-data-world/>
- Capra, F., & Sempau, D. (1998). *La trama de la vida*. Barcelona: Anagrama.
- Casanovas, P. (1998). Ensayo sobre la bondad: Guía práctica no sólo para la investigación de campo. In *La joven sociología jurídica en España: aportaciones para una consolidación* (pp. 27-68). The International Institute for the Sociology of Law.
- Curran, J. (2013). Big Data or “ Big Ethnographic Data ”? Positioning Big Data within the ethnographic space. In *American Anthropological Association, EIPIC Proceedings* (pp. 62–73).
- Denzin, Norman K. and Yvonna S. Lincoln. 2000. “Introduction: The Discipline and Practice of Qualitative Research.” Pp. 1–29 in *Handbook of Qualitative Research*, 2d ed., edited by Norman K. Denzin and Yvonna S Lincoln. Thousand Oaks, CA: Sage
- Evans-Pritchard, E. E. (1940). *The nuer* (Vol. 940). Clarendon: Oxford.
- Fallon, K., Davis, U. C., & Isbn, M. Z. (2014). Streams of the Self : the Instagram Feed As Narrative Autobiography, 54–60.
- Felt, M. (2016). Social media and the social sciences : How researchers employ Big Data analytics, (June), 1–15.
- Ferrándiz, F. (2011). *Etnografías contemporáneas. Anclajes, métodos y claves para el futuro*. Barcelona: Anthropos.
- Ford, H. (2014). Big Data and Small: Collaborations between ethnographers and data scientists. *Big Data & Society*, 1(2), 1-3.

- 
- Geertz, C. (1991). La interpretación de las culturas. Descripción densa: hacia una teoría interpretativa de la cultura. Barcelona, Ed. GEDISA.
- Gitelman, L. (2013). Raw data is an oxymoron. MIT Press.
- Gómez, E., & Thornham, H. (2015). Selfies beyond self-representation: the (theoretical) f(r)ictions of a practice. *Journal of Aesthetics & Culture*, 7.
- Gunthert, A. (2015) "La consécration du selfie. Une histoire culturelle", *Etudes photographiques*, n° 32, Spring 2015, <http://etudesphotographiques.revues.org/3495>.
- Hsu, W. (2012) On Digital Ethnography, What do computers have to do with ethnography? (1 of 4) <http://ethnographymatters.net/blog/2012/10/27/on-digital-ethnography-part-one-what-do-computers-have-to-do-with-ethnography/>
- Hsu, W. (2013) On Digital Ethnography, magnifying the materiality of culture (3 of 4) <http://ethnographymatters.net/blog/2013/01/25/on-digital-ethnography-magnifying-the-materiality-of-culture-part-3-of-4/>
- Hsu, W. (2014). Digital ethnography toward augmented empiricism: A new methodological framework. *Journal of Digital Humanities*, 3(1), 3-1.
- Kirschenbaum, M. (2014). What Is "Digital Humanities," and Why Are They Saying Such Terrible Things about It?. *differences*, 25(1), 46-63.
- Kitchin, R. (2014). Big Data , new epistemologies and paradigm shifts. *Big Data & Society*, (June), 1–12.
- Losh, E. (2013). Beyond Biometrics: Feminist Media Theory Looks at Selficity, Manuscript, Academia.edu.
- Malinowski, B. (1975). Argonautas del Pacífico occidental. Ediciones Península: Barcelona.
- Manovich, L. (2007). Cultural analytics: Analysis and visualization of large cultural data sets. *Manovich. net*, 30. <http://lab.softwarestudies.com/p/cultural-analytics.html>
- Manovich, L. (2011). *Trending: The Promises and the Challenges of Big Social Data* (pp. 1–17).
- Markham, A. (2013). Undermining 'data': A critical examination of a core term in scientific inquiry. *First Monday*, 18(10). doi:10.5210/fm.v18i10.4868

- 
- Nova, N. (2013). Ethnomining and the combination of qualitative & quantitative data. <http://ethnographymatters.net/blog/2013/04/02/april-2013-ethnomining-and-the-combination-of-qualitative-quantitative-data/>
- Pink, S., A. Markham, Y. Akama, E. Gómez Cruz, P. Lacasa, M. Poblet, S. Sumartojo (2016) DATA ETHNOGRAPHIES (2): ethics and data futures. Available online at <https://dataethnographies.com/paper-ii-ethics-and-data-futures/>
- Pink, S., D. Lupton, M. Berg, P. Dourish, A. Dyer, V. Fors, E. Gómez Cruz, H. Horst, P. Lacasa, John P., S. Sumartojo, and E. Witkowski (2016) DATA ETHNOGRAPHIES (1): personal data in an uncertain world. Available online at <https://dataethnographies.com/paper-i-personal-data-in-an-uncertain-world/>.
- Pink, S., E. Ardevol and D. Lanzeni (2016) 'Digital Materiality: configuring a field of anthropology/design?' in S. Pink, E. Ardevol and D. Lanzeni (eds) Digital Materialities: anthropology and design. Oxford: Bloomsbury
- Pink, S., Mackley, K. L., Mitchell, V., Hanratty, M., Escobar-Tello, C., Bhamra, T., & Morosanu, R. (2013). Applying the lens of sensory ethnography to sustainable HCI. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 20(4), 25.
- Rouch, J. (1995). El hombre y la cámara. En Ardévol, E. Pérez-Tolón, L. *Imagen y cultura. Perspectivas del cine etnográfico* (pp. 95-121). Diputación Provincial de Granada.
- San Cornelio, G. (2015). ¿Nos hacemos un selfie ? reflexiones y aproximaciones académicas al fenómeno. *Comme-In*, 1–10.
- Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Sage Publications, Inc.
- Taylor, S.J. y Bogdan, R. (1987). *Introducción a los métodos cualitativos de investigación. La búsqueda de significados*. PAIDOS: Buenos Aires
- Wang, T. (2013). big-data-needs-thick-data/. Retrieved from <http://ethnographymatters.net/blog/2013/05/13/big-data-needs-thick-data/>
- Zwitter, A. (2014). Big Data ethics, (December), 1–6.

