

## Datificación. Aportes para una discusión filosófica sobre la informática y sus repercusiones sociales

*Datafication. Contributions for a Philosophical Discussion on Computer Science and Its Social Repercussion*

Jeremías CAMINO

Universidad Nacional de La Plata, Argentina  
jereasecas@gmail.com

Recibido: 27/12/2021. Revisado: 17/02/2022. Aceptado: 16/03/2022

### Resumen

El siguiente artículo se propone establecer discusiones filosóficas sobre la datificación, basando sus reflexiones en conceptos elucubrados a partir de la ciencia de la informática. Se pretende enriquecer las consideraciones habidas en la bibliografía, respecto de los fenómenos informáticos a que da lugar la datificación, como el Big Data, el concepto de algoritmo, de base de datos y de usuario, mientras que se evalúa críticamente las apreciaciones comunes respecto de lo que es el dato, de la relación dato-mundo, de la vinculación entre la lógica y la informática, y de la datificación como nuevo dispositivo de control.

**Palabras clave:** dato; big data; algoritmo; virtualidad; usuario.

### Abstract

The purpose of the following article is to establish philosophical discussions about datafication, basing its reflections on concepts developed from computer science. It is intended to enrich the considerations in the bibliography, regarding the computer phenomena that datafication gives rise to, such as Big Data, the concept of algorithm, database and user, while critically evaluating the common appreciations regarding what which is the data, of the data-world relationship, of the link between logic and computer science, and of datafication as a new control device.

**Keyword:** data; big data; algorithm; virtuality; user.

## 1. Presentación general del tema de este artículo y sus problemas

En las últimas cuatro décadas, como es sabido, se ha instalado lo que se conoce como tecnología de la información, cuya definición general puede estimarse como todo aquello que permite y da soporte “a la construcción y operación de los sistemas de información” (Cohen Karen y Asín Lares, 2009, 8). Si bien el concepto más relevante de esta definición es el de información, sin embargo, en cuanto nos preguntamos a su vez por cómo entender este concepto, resulta que la información es, siguiendo a estos mismos autores, un dato o conjunto de datos dentro de un contexto comunicacional —es decir, que tiene un “significado para alguien” (Cohen Karen y Asín Lares, 2009, 3)—. Justamente, lo peculiar de las últimas cuatro décadas (lo cual se ha ido notando especialmente en las últimas dos) es un incremento inmenso de la cantidad de datos producidos, criterio que ha sido utilizado para determinar el enorme impacto que ha tenido la tecnología de la información en las sociedades<sup>1</sup>. Pues, obviamente, ese aumento solo podía deberse a una mayor presencia de los datos en las actividades humanas. En otras palabras, lo que ha ocurrido (y aún sigue ocurriendo) fue un proceso de *datificación*.

Ahora bien, este proceso ha dado lugar a ciertos fenómenos que han llamado la atención de numerosos estudiosos y pensadores, por la importancia adquirida en la vida de las sociedades. Nos referimos, entre otros, al fenómeno del Big Data (y la “abundancia” expectante que genera junto con su complejidad de análisis), que ha invitado a muchos soñadores a imaginar diseños de ciudades basados en la producción continua de datos. Otros autores, como Anderson (2008), han planteado que la datificación implicará, finalmente, el desplazamiento de la ciencia tal como la conocemos por una ciencia de los datos, en la cual el mecanismo de generar hipótesis es suplantado por el de mera correlación de datos. Y autores como Mayer-Schönberg y Cukier, sostienen (2013) que la tecnología de la información es la nueva revolución tecnológica. Por otro lado, es también reconocida en la bibliografía y en las percepciones y reflexiones comunes, el sostener que la enorme producción de datos implica un más exhaustivo control de los individuos y las sociedades (van Dijk, 2014), valoraciones en las que juega un papel relevante la presencia invisible (y no del todo entendida) de los algoritmos.

Ahora bien, creemos que en estos estudios valiosos se plantean conclusiones basadas, en general, en los efectos más perceptibles del asunto, como es la contabilidad y cuantificación de datos, la inmediatez de los procesos datificados, la

---

<sup>1</sup> La CEPAL (2015, 11) recoge, entre otras informaciones, que, para el 2014, el 50% de la población era un usuario de internet (p. 11), que el tráfico por IP era de 60 000 exabytes al mes (p. 17), que el usuario promedio contaba con sesenta aplicaciones (p. 20), que, entre el 2005 y el 2013, la mayor datificación de la cadena de valor de las industrias ha modificado la organización productiva (p. 22), y que, en 2014, las ventas de comercio electrónico representaba el 2% del PIB mundial, impulsadas por la datificación.

delegación de tareas humanas en la operatividad de los sistemas de datos, la relativa facilidad de obtener y operar con datos, la eficacia organizacional, etc., sin abordar la cuestión desde lo que sea un dato y qué sea la datificación.

Por ello, este trabajo se propone contribuir a la discusión filosófica del asunto, cuyas elucubraciones tengan en consideración, principalmente, la ciencia de la informática (siempre abordada filosóficamente). Lo cual se llevará a cabo de la siguiente manera. Primero, obtendremos el concepto de dato y de datificación, enfocándonos en los elementos informáticos pertinentes para las discusiones posteriores, como es el de base de datos relacional y el de algoritmo. Esta disquisición nos impondrá reflexionar sobre la insalvable diferencia entre la informática y la lógica (debido a que podría creerse que, elementos que comparten, las hace convergentes), así como problematizar críticamente la idea común de que el dato representa una porción del mundo (se demostrará que la datificación implica pensar la relación dato-mundo de otra manera). En segundo lugar, discutiremos sobre el impacto que la datificación ha tenido en la vida de las sociedades, enfocándonos muy particularmente en el concepto de usuario, al cual entendemos como la forma datificada de los individuos humanos. Veremos que en ese concepto convergen paradigmáticamente los elementos informáticos que habremos desarrollado, cuyo entendimiento nos permitirá conceptualizar la relación triádica entre el individuo, el consumo y producción de datos, y el sistema de datos en donde esto se realiza (sistema detrás del cual hay una empresa que está interesada en que ello se efectúe). Así, discutiremos críticamente la idea común de que estamos en presencia de una nueva forma del “Big Brother” orwelliano o de un Panóptico digital, ya que los procesos de datificación habilitan, antes que nada, dinámicas sociales fundadas, no tanto en la sujeción o sometimiento propio de las sociedades disciplinarias, sino en la manipulación (más acorde al producto de la datificación). Finalmente, nos adentraremos en el Big Data. Todo el desarrollo previo nos permitirá establecer qué es este fenómeno informático, estableciendo una discusión crítica sobre qué esperar de ello, ya que, aunque el Big Data surge debido al rompimiento de ciertos condicionamientos o limitaciones de los procesos de datificación habituales (instalando la novedad de producción de datos a partir exclusivamente de datos), sin embargo, se lo hace al precio de una especie de indeterminación, que lo lleva a un estado difuso del cual, según parece, solo puede tenerse un aprovechamiento parcial.

## 2. Aclaración histórica sobre la ciencia de la información

Antes de comenzar, sería conveniente hacer un breve comentario sobre esta ciencia de la informática en que nos basaremos a continuación. En primer lugar, debe considerarse que la creación de datos que hacen a la información a que hemos referido, depende de un proceso técnico cuyas herramientas de producción fueron creadas en el s. XX. El “sistema tecnológico” (como lo denomina Castells, 2000, 91) que da sustento a la datificación actual, tiene su campo de desarrollo,

particularmente, en la micro ingeniería y la utilización del silicio. Por estos dos, los computadores creados durante la Segunda Guerra Mundial tienen una transformación crucial, porque permiten la construcción de computadores más pequeños, mejoran drásticamente el almacenamiento y transmisión de datos, y, consecuentemente, el poder de cálculo. En segundo lugar, al desarrollo de esta tecnología han contribuido diversas empresas e incluso estados nacionales (a través de sus institutos de investigación), cuya expansión comercial se da a partir de la década del setenta (Castells, 2000). Por lo tanto, y a pesar de que aparenta estar integrado, el desarrollo de estas tecnologías estuvo esparcida en diferentes actores privados, lo que supone una diversificación en cuanto a su conceptualización, a su teorización. En otras palabras, al menos hasta finales del siglo pasado, no había una teoría única o ciencia que sustentara conceptualmente el desarrollo tecnológico, sino, más bien, concepciones independientes solo prácticamente compatibles. Aunque esto no es atípico en el desenvolvimiento de la ciencia moderna (como lo demuestra su historia desde Galileo Galilei en adelante), sin embargo, no hay una teoría equivalente a la que fue, por ejemplo, los *Philosophiae aturalis principia mathematica* de Newton para la física moderna, o *El origen de las especies* de Darwin para la biología moderna.

No obstante, la obra *Fundamentos de sistemas de bases de datos* de R. Elmasri, R. y S. B. Navathe, publicada en su idioma original en el año 2006, si bien está pensada estrictamente para explicar los “conceptos fundamentales” de las bases de datos utilizadas en actividades humanas que requieren de un registro datificado (como ministerios, empresas privadas, universidades, escuelas, etc.), también contiene las bases científicamente formuladas para una ciencia informática que resulta clara y coherente. Aun así, podría preguntarse ¿por qué comenzar con una base de datos - que es el elemento central en dicha obra? Lo que en nuestra introducción hemos denominado como “tecnología de la información”, involucra, en sentido general, un “equipo computacional”, un “recurso humano”, “información fuente”, “programas”, “telecomunicaciones” y “procedimientos” (Cohen Karen y Asín Lares, 2009, 6). Esto se debe a que, históricamente, siguiendo a Castells, las primeras fuentes de *creación de datos* estuvieron ligadas a emprendimientos de diversa índole (como organización de fábricas o instituciones públicas) así como también a la telecomunicación (como lo fue inicialmente el proyecto militar ARPA de Estados Unidos), que tuvieron como herramienta central al computador u ordenador (los nombres no son azarosos) que, justamente, creara, procesara, almacenara, recibiera y transmitiera datos. El computador fue el instrumento que, en gran medida, intermedió en el proceso de datificación; y, para cumplir con las funciones requeridas, era preciso una base de datos. En otras palabras, la datificación emerge estando instrumentalizada por el computador y la base de datos. Actualmente, no es necesario que toda actividad datificada esté mediada por un computador (pues, por caso, existe el celular); tampoco es necesario que todo artefacto del estilo requiera hacer todas las funciones recién

mencionadas (podrían crearse datos de existencia fugaz que no precisen almacenamiento). Pero, sin una base de datos (y lo que ella conlleva) y un instrumento apropiado para ello, difícilmente pueda mantenerse la datificación actual.

La datificación, entonces, ocurre junto con el proceso de construcción de una base de datos. Este último proceso está ligado específicamente a los fines para los cuales se la requiere; por lo tanto, como veremos, los elementos que la constituyen también resultarán específicos. No obstante, hay conceptos y elementos generales de cualquier base de datos que nos permitirán establecer las discusiones filosóficas que pretendemos, y, por lo tanto, serán los que desarrollemos con cierta profundidad.

### 3. Lo que es un dato y la relación con el mundo

#### 3.1 El concepto de dato

Siguiendo a Elmasri-Navathe (2007), una base de datos informática se puede ubicar dentro del género de las colecciones ordenadas de datos, con fines de registro, almacenamiento y consulta. Lo que la diferencia específicamente del resto es el modo en que esos fines se llevan a cabo. Esto se torna relevante por cuanto determina definitivamente la esencia del dato. En informática, existen diversas formas de datificar una actividad humana y las cosas con las que ésta se maneja; aquí nos centraremos en la de tipo relacional, que surgió a partir del trabajo realizado por E. F. Codd (para la empresa IBM), en la década del 70, por su ya reconocida importancia. Si bien existen versiones ampliadas de este mismo sistema, y otras alternativas (como el que Elmasri y Navathe denominan entidad/relación), su desarrollo nos permitirá entender lo que es un dato.

Para construir una base de datos, en una primera instancia, el constructor (o programador) debe seleccionar las notas características y destacables de la actividad humana y de las cosas con las que esta actividad trata. Son lo que Elmasri y Navathe llaman las “cualidades” y “relaciones”. Por ejemplo, en una empresa comercial, una cualidad sería el ser empleado, el ser jefe de una sección o la gerencia, el tipo de materias primas y la cantidad en disponibilidad, etc.; mientras que una relación sería un proyecto de trabajo, una sección de la empresa, las funciones a cargo, o los empleados a cargo, etc. Cada una de las cualidades y de las relaciones, subsiguientemente, es estructurada en lo que se conoce como un *esquema relacional*. Tal vez, podría sospecharse que un esquema relacional es un predicado lógico; sin embargo, esto no es cierto. Aún más, la lógica se diferencia insalvablemente de la informática por dos motivos que examinaremos a continuación, para establecer una correcta consideración de lo que es la datificación.

En primer lugar, se recordará que la definición contemporánea de predicado lógico es, más allá de posibles sofisticaciones ocurridas durante el s. XX, la que impulsara Frege en su *Conceptografía*, esto es (según sus términos), una “función”

cuyo “argumento” es el conjunto de cosas o individuos de los cuales se predica (o podría ser aplicable), en el sentido de que hacen verdadera la oración “de esta cosa se predica tal propiedad”<sup>2</sup>. Diferentemente, un esquema relacional está conformado mediante lo que técnicamente se denominan *atributos*, los cuales representan tanto a las cualidades como a las relaciones. Cada atributo delimita, a su vez, un conjunto de “espacios vacíos” que se pueden rellenar y se denominan técnicamente *instancias*. Así, se tiene la imagen habitual (promovida por el propio Codd) de un esquema relacional como una tabla, donde el conjunto de instancias bajo un atributo conforma una columna, y el conjunto de instancias agrupadas por una “fila” se denomina técnicamente *tupla*<sup>3</sup>.

También, sería incorrecto pensar que los esquemas relacionales son como los predicados de predicados que trata la lógica de segundo orden. En el mejor de los casos, estos predicados de segundo orden siguen estando pensados como inicialmente los había planteado Frege, es decir, como cosas insaturadas. Un esquema relacional, en términos comparativos (quizás no exactos), es algo que cada vez está “completo”, saturado.

<sup>2</sup> Frege (1972, 14) había escrito: “Si en una expresión cuyo contenido no necesita ser juzgable, aparece un símbolo simple o compuesto en uno o más lugares, y si lo pensamos como reemplazable en todos o en algunos de estos lugares por algo distinto, pero siempre por lo mismo, entonces a la parte de la expresión que aparece sin cambio la llamamos función y a la parte reemplazable, su argumento”. Lo que Frege (1984) entiende por función y argumento lo explica claramente en su ponencia del 9 de enero de 1891, “Función y concepto”, donde presenta a los predicados como los elementos no-saturados de una oración, es decir, que tienen uno o más lugares vacíos para ser rellenados con los argumentos. Un siglo después, GAMUT (2002, 69 y ss.) recoge las ideas principales de la propuesta fregeana, al hablar de constantes que refieren a un individuo u objeto, y que están “fijados” a las letras de predicado que refieren a propiedades que los individuos u objetos tienen, formalismo que, a su vez, requiere del “diccionario de traducción” y el universo de discurso.

<sup>3</sup> Un ejemplo tomado de Elmasri y Navathe (2007, 125 y ss.). En el registro de una universidad, una base de datos de sus estudiantes puede construirse, en parte, sobre el esquema relacional ESTUDIANTE, conformado por los atributos: Nombre, DNI, Teléfono Particular, Dirección, Carrera, Cantidad de Materias Aprobadas, Promedio. Obsérvese que el esquema relacional contiene valores o instancias cuyo “tipo de dato” tiene caracteres numéricos y también alfabéticos. Una tupla es el conjunto formado por una “fila” de instancias (a tal Nombre corresponde tal DNI, tal Teléfono Particular, etc.), por lo que habrá tantas tuplas como estudiantes. De esta forma, puede pensarse a una tupla (según Elmasri-Navathe) como describiendo, en los términos de los atributos, a un estudiante. Un punto importante a destacar es que, en cualquier momento, el esquema relacional está “completo” (dicho en una terminología comparativa). Nótese, además, que, para esta universidad, también puede ser importante para su organización, tener especificado el código del departamento especial que le corresponde a cada estudiante según su carrera. Por lo tanto, podría existir el esquema relacional ESPECIALIDAD conformado por los atributos: DNI, Código de Departamento. Este esquema estructura una propiedad lógicamente unaria y otra lógicamente diádica.

En segundo lugar, todas las computaciones realizables en una base de datos están fundamentadas en el álgebra relacional<sup>4</sup>, el cual opera *directa y únicamente* con esquemas relacionales. El álgebra relacional fue creado utilizando algunos conceptos de la teoría matemática de conjuntos junto con desarrollos específicos para las bases de datos (Codd, 1970; Elmasri y Navathe, 2007, 146). Tales computaciones no son las operaciones de la lógica, ni en sentido semántico (como el cálculo de tablas de verdad) ni sintáctico (como el sistema de deducción natural de Gentzen)<sup>5</sup>. Pues, no están pensadas para evaluar argumentos sino para generar, óptimamente, las interacciones de un usuario con la base de datos, las cuales son: consulta, agregación, eliminación y modificación de datos. Estas se llevan a cabo mediante computaciones que están determinadas por una o varias operaciones del álgebra relacional, encadenadas de una manera específica para garantizar siempre un resultado. La secuencia de operaciones algebraicas y cuáles sean las que deben utilizarse *depende* del tipo de datos de la base de datos y del tipo de consultas que se realizará. Como hay un orden de ejecución programado, a tales computaciones se las conoce como algoritmos. Esto implica que no cualquier operación algebraica ni cualquier modo de ejecutarlas puede ser realizada, lo cual hace a los condicionamientos o límites específicos de la cantidad posible de resultados y de su tipo<sup>6</sup>. Es fundamental señalar, además, que todas las operaciones algebraicas, y, por lo tanto, los algoritmos, parten de esquemas relacionales para llegar a un resultado que, o bien es una instancia, o bien un conjunto de instancias (ordenadas en tuplas, columnas o alguna combinación de una u otra), o bien un nuevo esquema relacional; es decir, que cualquier resultado está en un esquema relacional o es un esquema relacional. Es el aspecto recursivo de una base de datos.

Si el esquema relacional y las operaciones relacionales conforman fundamentalmente la estructura elemental de una base de datos, entonces, ¿qué es un dato? Para alcanzar la respuesta, procederemos analizando los candidatos posibles de todos los elementos que enumeramos anteriormente, para generar una idea cada

<sup>4</sup> Elmasri y Navathe (2007, 145) señalan una diferencia técnicamente importante entre álgebra y cálculo. La primera contiene los fundamentos de las operaciones realizables con los elementos que son los esquemas relacionales, mientras que la segunda es el conjunto de expresiones codificadas en un lenguaje acorde al sistema de base de datos (es el lenguaje DDL: ver el apartado 4).

<sup>5</sup> Nos basamos en la distinción que establece GAMUT (2002, 122 y ss.), centrada en el concepto de argumento y validez.

<sup>6</sup> ¿Cuáles son estas relaciones? Codd en su artículo (y posteriores ampliaciones) las especificó, y son las mismas que tratan y explican detalladamente Elmasri y Navathe (2007, Cap. 6). Las más elementales de estas son: la operación selección (que selecciona la, o las, instancias agrupadas en las tuplas de un esquema relacional), proyección (que selecciona la o las instancias agrupadas en columna de un esquema relacional), unión (que junta, de dos esquemas relacionales, las instancias que están en uno, en otro y en ambos), la intersección (que junta las instancias que están en ambos esquemas a la vez), la diferenciación (que junta las instancias que están en un esquema pero no en otro), el producto cartesiano (que combina todas las instancias de dos esquemas), y la concatenación (que encadena el producto cartesiano con la selección).

vez más clara. Así, si se propusiera que el dato es una instancia, se tendría una definición deficiente, porque la instancia es lo que es debido, en parte, al atributo al que corresponde (un número o un nombre, por sí solos, no expresan nada). Si se propusiera que el dato es uno de los atributos, también sería insuficiente. Por un lado, un atributo aislado no es expresivo de algo. Por otro lado, un atributo y una de, o todas, sus instancias, no tiene demasiado sentido; piénsese en cuál sería el significado (y la utilidad) de tener el conjunto de, por ejemplo, todas las instancias del atributo 'DNI' de una parte de la población. La instancia (o cualquier conjunto de estas) y el atributo comienzan a tener sentido cuando son considerados como parte del esquema relacional al que pertenecen; así, siguiendo el ejemplo anterior, las instancias 'DNI' se agruparían bajo el esquema, v.g., JUBILADO, que, a su vez, tiene un sentido dentro del sistema estatal de pensiones jubilatorias. Al hacer esto, para hablar de dato resultará necesario considerar al resto de los atributos de ese esquema (que contendrá posiblemente atributos como 'nombre', 'edad', 'dirección', etc.). No obstante, esto también será insuficiente, ya que habrá otros esquemas relacionales. Por ejemplo, podría ser provechoso disponer de los individuos cuyo 'DNI' termine en una cifra par y pertenezcan a la provincia tal o cual, lo cual combinaría dos esquemas relacionales (JUBILADO y PROVINCIA), y que, así, constituirían la base de datos de los pensionados.

Entonces, se tiene la siguiente situación. Una instancia (v.g., el número de DNI) es, efectivamente, un dato si está integrado en un sistema de datos. Pero también es un dato el conjunto de instancias que demarca un atributo (o un conjunto de atributos), así como también una tupla (o conjunto de tuplas), tanto como un esquema relacional íntegro (o conjunto de esquemas). ¿Cuál es la razón de que sea así? La respuesta se debe considerar en dos aspectos equivalentes. Por un lado, cada instancia o conjunto de instancias (ordenadas en tuplas y/o columnas y/o esquemas enteros) puede entenderse como el resultado derivado de las operaciones del álgebra relacional. Por otro lado, la totalidad de las instancias es entendible como la junción de todos los esquemas relacionales habidos en la base de datos. Por lo tanto, el conjunto de todos los resultados operacionales posibles es un conjunto de datos, el cual, a la vez, es también un dato, así como el conjunto de todos los esquemas relacionales es un conjunto de datos, el cual también es un dato<sup>7</sup>.

De lo anterior, se deriva que: 1. El dato es un término de masa (como el término 'agua'), es decir, cada parte de un dato es también un dato, 2. El dato es un

<sup>7</sup> Este doble aspecto se debe a dos factores concurrentes de una base de datos (que, de hecho, la caracterizan). Primero, que toda el álgebra relacional opera con los datos almacenados, estructurados en el esquema relacional. Segundo, que cualquier resultado siempre y necesariamente estará dentro de un esquema relacional. Estos aspectos también quedan expresados así: las operaciones algebraicas relacionales operan con esquemas relacionales y sus resultados están dentro de estos (en jerga matemática, rango y dominio coinciden). Es el carácter sistemático de la base de datos (su nombre técnico descriptivo es DBMS, *database management system*).



concepto perfectamente definido aunque complejo, porque cada parte del dato está virtualmente entrelazada con una multiplicidad de partes del dato a través de las operaciones algebraicas, característica que explica la percepción usual del entramado de datos. Este entrelazamiento virtual, sin embargo, no es infinito (en ningún sentido), porque todos los enlaces virtuales están limitados por aquellas condiciones específicas que mencionamos (los algoritmos) que define el programador y porque parte de una cantidad siempre finita de datos. Estas dos conclusiones serán relevantes en el apartado de las discusiones, en que trataremos el concepto de usuario y el fenómeno del Big Data.

### 3.2 Datificación y representación

Ahora bien, ¿se deriva de lo anterior que el dato es *necesariamente* algo representativo de una parte del mundo? Como demostraremos a continuación, la respuesta es sí y no. Quizás, lo más importante de destacar (y anticipar) es que el dato, por sí mismo (=para ser y existir), no precisa de ser representativo del mundo, porque, a la larga, el efecto que la datificación genera es la indistinción mundo-dato. Para demostrarlo, hay que comenzar estableciendo qué entender por ‘representación’, ya que este concepto se dice de muchas maneras. Comenzaremos suponiendo que es el significado, esto es, que el dato es representativo en tanto que significa algo del mundo. Cabe añadir que, debido a que la producción de un dato (y consecuentemente su definición) es algo técnico (preciso y exacto), como lo es el lenguaje artificial de la lógica o de la matemática, no necesitaremos entrar en las peripecias argumentales de aquellos lógicos que buscaron entender el significado del “lenguaje natural” durante el siglo pasado<sup>8</sup>.

Nos preguntamos algo sencillo: ¿en qué sentido el dato *significa* algo del mundo? Caben dos alternativas. Una es comprender la significación en sentido extensional, como cuando el concepto “perro” significa (o denota) al conjunto de los animales de los cuales se dice verdaderamente que son perros. Paralelamente, deberíamos encontrar esa relación denotativa en el dato, y ver qué elemento del mismo se comporta extensionalmente como el concepto “perro”. De nuestra exposición anterior, los candidatos más obvios (y quizás los únicos que puedan postularse) son los atributos y sus instancias. Ahora bien, es cierto que, para que una empresa o una dependencia estatal *funcione* como se espera, resulta importante el que las instancias delimitadas por los atributos denoten a los entes del mundo involucrados en la actividad humana. Sin embargo, hemos visto que, ni la instancia ni el atributo, por sí solos, tienen sentido, sino en tanto son parte de un esquema relacional dentro del sistema de la base de datos. Por lo tanto, las instancias y los atributos siempre son lo que son *para* un esquema. Si ellas refieren o no a un estado de hecho, no es entitativamente necesario. Un ejemplo

<sup>8</sup> Para una excelente documentación del debate ocurrido desde principios del s. XX en adelante en la tradición analítica, ver el libro de Valdés Villanueva *La búsqueda del significado*.

claro es que puede fabricarse una base de datos de cosas realmente inexistentes, y funcionar perfectamente como una base de datos usual, con usuarios interactuando como si existiera aquello que estaría datificado (veremos otros ejemplos concretos en breve).

El otro sentido de significación es el intensional (del latín *intensio*), es decir, la connotación. Pero, si con ello se pretende decir que el dato tiene una definición, como, p.ej., el triángulo es una figura de tres ángulos, ¿cuál sería el paralelo en el dato? Para comenzar, la propuesta más inmediata sería el conjunto de atributos de un esquema relacional. Sin embargo, siguiendo nuestra exposición anterior, esta propuesta sería parcial, ya que un dato bien podría ser un esquema relacional, pero también combinaciones de atributos (y las instancias demarcadas por cada uno de estos) que los algoritmos permitan. Una mejor propuesta, más general, sería el conjunto de atributos lícitamente combinados. De esta forma, esa combinación algorítmicamente permitida estaría conformada por los atributos, que tendrían el lugar de “figura” y “tres ángulos” de la definición “triángulo”. Ahora bien, es cierto que, como antes, será importante para el funcionamiento adecuado de una empresa o dependencia estatal, el que las combinaciones permitidas estructuren coherente y “verdaderamente” atributos. Sin embargo, por otro lado, ello no es necesario para que el dato sea en tanto tal, pues, en el dato en sí, solo importa la estructuración por y para la computación: no importa *qué* es lo que se estructure, sino *que* esté estructurado.

Ciertamente, la consecuencia más inmediatamente visible es que el dato mantiene un grado de independencia respecto del mundo. Sin embargo, hay algo más destacable en esto que se puede captar a partir de los siguientes tres ejemplos. El primero es el de los mapas de los aparatos GPS. Estos son, indudablemente, una base de datos. El usuario interactúa con el mismo solicitando que lo oriente hacia un destino deseado; siguiendo sus indicaciones, llega sin haberse preocupado en lo más mínimo de la ruta que efectivamente recorrió. Para el usuario, ver el GPS o ver la ruta es indistinto, no tanto por la condición epistémica de que así lo crea (o no), sino porque el GPS es un instrumento que se presenta en relación con la ruta efectiva como siendo lo mismo. El segundo ejemplo es el que expone Morozov (2011, Cap. 1), al relatar cómo algunas sociedades europeas vivieron la “supuesta” revolución iraní del 2009, que recuerda a lo que Baudrillard había anticipado a fines de los noventa: hacer la revolución en “tiempo real” (2000b). Este fenómeno fue mediado por la divulgación de la red social Twitter, que es, estrictamente dicho, una base de datos (como todas las redes sociales, como discutiremos en la siguiente sección). Y finalmente, el ejemplo por antonomasia, son los juegos de realidad virtual, y muy particularmente el Pokemon Go. Se recordará que este consistía en que el jugador “encontrara por la ciudad”, guiado por la aplicación correspondiente, a los personajes de la serie animada “Pokemon”. Se podía ver a niños y adolescentes compenetrados en esa tarea, guiados

por la aplicación del celular que, a la vez que daba la locación precisa del usuario, también daba la del personaje buscado. En todos los casos, la conclusión es exactamente la misma que la del ejemplo del GPS.

La relación entre la base de datos y el mundo no es la misma que la de una pintura como “La escuela de Atenas” de Rafael de Urbino, en la cual interviene la imaginación en un espacio donde se da un juego entre la verdad y la falsedad, y se transmite un significado simbólico. La anterior discusión sobre la representatividad del dato en tanto significado concluye que la datificación elimina la denotación y la connotación, en el sentido al que apuntan los ejemplos anteriores. Esto es, que el producto de la datificación está en algo así como una “interrelación” con el mundo, por lo que tiende a no tener sentido la pregunta de si hay, o qué es, aquello de lo que los datos “hablan”, sino que estos se establecen como siendo eso mismo que “significarían”, de manera indiferenciada del mundo. Es una relación semejante a la que instala la escultura hiperrealista (o el arte hiperrealista en general), que no solo es una reproducción en una escala diferente del objeto esculpido, sino que se muestra como siendo eso mismo, como si no hubiera distinción esencial. En este caso, se rompe el juego entre la verdad y la falsedad, y el significado se torna literal. Este paralelo nos permite pensar que la datificación recrea algún aspecto del mundo. Sin embargo, dicho proceso tiende finalmente a que no exista diferencia entre ver el mundo y ver los datos.

## 4. Discusiones

### 4.1 Tecnología de la información

A continuación, discutiremos algunos de los aspectos de la repercusión social de la datificación. En este apartado, queremos analizar, desde el basamento anterior, algunas de las actividades digitales más habituales, como son las redes sociales y la navegación en internet a través de algunos buscadores populares. El elemento en que nos centraremos para ello es el de usuario, al que, anticipando la discusión, definimos como la forma datificada de un individuo que le interesa ingresar a las bases de datos. En el usuario convergen muchas de las cuestiones anteriormente desarrolladas, cosa que nos permitirá tener una perspectiva anclada en la ciencia de la informática para considerar aquellas aplicaciones de la datificación de uso diario.

Ahora bien, para lograr el objetivo de una forma lo suficientemente clara, será preciso introducir un ulterior concepto técnico. Establecimos que una base de datos se construye con esquemas relacionales y las operaciones del álgebra relacional o algoritmos. Estos permiten al usuario interactuar con la base de datos. A la vez, se requiere que todo ese entramado de esquemas relacionales y algoritmos quede expresado en un lenguaje único, común a ambos (y otros aspectos no abordados aquí), que, al mismo tiempo, permita una visualización más amena de la base de datos. El nombre técnico de tal lenguaje es DDL (*data definition*

*language*). La breve historia de aplicaciones y desarrollos del modelo y su DDL después del artículo de Codd, generó una de las variadas formas de crear este lenguaje más utilizadas que es el SQL (*structure quering language*).

Entonces, nos preguntamos ¿qué es un usuario? Comenzaremos con una aproximación que complejizaremos para alcanzar su concepto. Un usuario es un espacio particular *en* la base de datos. En principio, decimos que se caracteriza por un perfil (que depende del lugar que le corresponda en la actividad general) y porque puede interactuar, esto es, puede consultar, agregar, modificar y eliminar datos. A este conjunto se le denomina en inglés *application*, que la literalidad ha traducido al español como aplicación, aunque el significado correcto es solicitud. Así, en principio, podemos decir que el usuario es aquello que decide qué operación llevar y quien la ejecuta.

Sin embargo, esta definición no está libre de ambigüedades. Véase que: *i*) las solicitudes se realizan mediante los algoritmos fundados en el álgebra relacional cuyos resultados, como se ha visto, son datos, y que *ii*) el dato está virtualmente entrelazado, dicho específicamente, que los datos ya están entrelazados entre sí según las formas que permiten los algoritmos particulares; por lo tanto, se debe concluir que, cada solicitud del usuario produce un dato, entendiendo por ‘producir’ el pasaje de lo virtual al hecho solicitado<sup>9</sup>.

Esto significa que el usuario está datificado, porque sus acciones son datos, son las “marcas” que va dejando en su interacción, y que quedan registradas en la base de datos. El usuario, pues, es un dato; la única y crucial diferenciación con cualquier otro dato es justamente que, el suyo, es el lugar de decisión y ejecución. Y, en tanto es un dato, estará sujeto a las caracterizaciones discutidas anteriormente. Cualquiera de sus decisiones y ejecuciones, y cualquier resultado que pueda esperar de ello, ya está determinado en la base de datos, porque, como ya hemos dicho, esas actividades se llevan a cabo por los algoritmos. Por lo tanto, el usuario está virtualmente determinado en su integridad. Esto nos lleva a preguntarnos cuál es el sentido de un usuario en una base de datos. Amén de la obvia respuesta de que el usuario es el individuo datificado para quien importa tener una base de datos, esta requiere, a su vez, de una cosa como un usuario para que se lleven a cabo las computaciones para las cuales fue diseñada. Por sí misma (y esto vale para cualquier sistema informático) la base de datos no puede hacer nada, sino que alguien, en tanto usuario, debe hacerla funcionar.

Cuando esto es modelado en una tecnología social, la datificación se presenta de diversas formas, una de las cuales son las redes sociales, tales como Facebook, Twitter, etc. Facebook, por ejemplo, es una base de datos cuyo lenguaje DDL es

<sup>9</sup> Hay que tener especial cuidado en no confundir virtualidad con potencia, en el sentido aristotélico del término. La definición de lo potencial (*dynamis*) que Aristóteles da en su Física (1995, 209a9-11) es la de aquello que aún no presencia cumplimentadamente (*entelécheia*). Lo virtual no comparte ninguna de esas características.

aquel que el usuario visualiza. Cuando el usuario “navega” y mueve la pantalla hacia abajo, “decide” y ejecuta la solicitud (*application*) de consultar continuamente datos; igualmente, cuando agrega contenido, realiza la solicitud de agregación de datos. Todas esas acciones se realizan determinadas invariablemente por el DDL de la red social, porque no hay otra forma de interactuar con la misma que no sea a través de sus algoritmos<sup>10</sup>. Esto tiene dos importantes implicaciones. Por un lado, cuando el usuario decide ejecutar una solicitud, está pretendiendo algo datificado. Así, por ejemplo, el agregar un “contenido” a la red social, trae consigo el amoldarlo al DDL de la misma. Si bien la red social se presenta inicialmente en tanto *herramienta* para autopublicitarse y para relacionarse con otros, genera una transformación, como toda herramienta, en la forma en que se dan esas acciones. Tal modificación depende, en este contexto, de cómo sea el DDL. Y ¿cómo es este molde? Eso varía según cada red social; Facebook, por ejemplo, permite comentarios, fotos, videos, o “reacciones” seleccionadas en una lista finita. Esto se determina en la construcción de dicha base de datos, y por lo tanto, depende de la intención de sus creadores. Lanier (2018) relata (y pretende acompañarlo con “explicaciones científicas” de) qué tipo de cosas se tenían en consideración al crear esas bases de datos desde Silicon Valley – como es, según dice, el conductismo. Como hemos visto, esta construcción ocurre mediante una selección de notas características de una actividad humana, que en el caso de las redes sociales tiende a ser una simplificación *en pos* de una optimización organizacional, como es la lista de reacciones de Facebook que manifiestan emociones datificadas (“me gusta”, “me enoja”, “me divierte”, etc.) y se pretende que medien las relaciones sociales. Aunque ciertamente sean formas de expresión de las emociones, conviven con otras formas no datificadas (que ocurren en el contacto presencial). Esto significa que hay una parte de la subjetividad y de las relaciones sociales que transmuta en lo que, en inglés, se designa con el término elocuente de *mining life*, minería de la *bio*-grafía (o la vida como una mina de datos).

A nuestro entender, esto prosigue en la línea de un fenómeno propio del postmodernismo, que se dio en llamar posthumanismo o transhumanismo. Utilizando la clasificación de Diéguez (2017, Cap. 1), podremos decir que, el posthumanismo de la datificación, corresponde al transhumanismo tecnocientífico. No hablamos aquí de la idealización de algunos entusiastas de la pervivencia humana en un software<sup>11</sup>, sino del modo ciborg de ser<sup>12</sup>. Así como el “cuerpo natural” se combina con el “cuerpo artificial” (maquinal), la subjetividad “natural” se combina con la subjetividad digital, en la misma medida en que la realidad lo hace con la virtualidad.

<sup>10</sup> De hecho, la “aplicación”, aquello que se descarga en el teléfono móvil o la computadora, es el software para ejecutar tales operaciones. De allí su nombre.

<sup>11</sup> Incentivados, sin dudas, por producciones cinematográficas como “Trascendens”.

<sup>12</sup> En el sentido en que lo trabajó artísticamente, por ejemplo, Stelius Arcadiou.

Por otro lado, cuando el usuario ejecuta una solicitud en una red social, el sistema trabaja únicamente con todos los datos que hasta ese momento han sido cargados, como si estuviera completo (carácter que no cambia, aunque pueda almacenarse más). Así, de acuerdo con la discusión del concepto de dato como término de masa, entendemos que cada vez ocurre que todos los datos están virtualmente entrelazados, están todos virtualmente dados, aunque parcialmente ejecutados por el usuario. Esto genera la apariencia de un control total por parte de la empresa dueña de la base de datos. Pero, esto habría que reflexionarlo un poco más detenidamente.

La datificación de una base de datos, como se puede anticipar, supone una construcción pormenorizada que implica una determinación íntegra de todo lo que la conforma y de todo lo que pueda y no hacerse. Esto significa que, aunque el usuario pueda decidir y ejecutar, todo lo que él es (=perfil e interacciones) ya está virtualmente determinado. Ello no es lo mismo que controlar, si es que por tal se entiende algo semejante a un “Big Brother” orwelliano o un panóptico benthamiano (como suele afirmarse en la bibliografía, van Dijk, 2014; Granroth, 2020). Esto lo ha desarrollado correctamente Byung-Chul Han en *Psicopolítica* (2014); en los apartados “Poder inteligente”, “El topo y la serpiente”, “Biopolítica” y “El dilema de Foucault”, discute acertadamente contra la idea de que la sociedad neoliberal (como él la llama), que no solo se caracteriza por la datificación (ver también “La sociedad de la información” – Han, 2013) sino que está producida en gran medida por tal proceso (Castell, 2000), sea una sociedad de control, de disciplinamiento. Por un lado, la lógica de este último tipo de aparataje es el de configurar un espacio determinado por limitaciones, o restricciones, expresadas a modo de deberes, obligaciones, derechos, o concepciones ideológicas, etc. Cada individuo debe someterse a estas condiciones, conducirse por esos carriles, a riesgo de ser castigado. Esto implica la tensión entre la necesidad del monitoreo continuo y la necesidad de la libertad (de una liberación de los controles). En cambio, la datificación, como se ha visto, requiere de la producción continua de datos que trae consigo, necesaria y simultáneamente, las solicitudes del usuario: *se precisa de las interacciones del usuario* (sin ello, como dijimos, las bases de datos no tienen sentido). No se requiere monitorear, sino, por sobre todo, crear usuarios y, a la vez, incentivar a la creación de más datos. “En lugar de hacer a los hombres *sumisos*, intenta hacerlos *dependientes*” (Han, 2014, 29). Han capta la existencia de un “poder inteligente” que “no opera de frente contra la voluntad de los sujetos sometidos, sino que dirige esa voluntad a su favor”. Y perspicazmente agrega que dicho poder no “impone ningún silencio. Al contrario: nos exige compartir, participar, comunicar nuestras opiniones, necesidades, deseos y preferencias; esto es, contar nuestra vida”, que es la *mining life* antedicha. Una vez hecha la datificación, es posible utilizar esos datos creados en pos de algún fin particular (como lo fue el caso de la Cambridge Analytica destapado en el 2018); pero, es un problema de manipulación, de utilización, que es diferente al del sometimiento o la sujeción (véase que, incluso, el proceso de manipulación

se presenta, en primera instancia, como algo claramente despersonalizado, como si fuera una cosa comercial<sup>13</sup>: el individuo se datifica en datos que se almacenan, y luego son utilizados para dirigir al *usuario* en una dirección preferencial, cuyo resultado efectivo será, nuevamente, un conjunto de datos)<sup>14</sup>.

Han expresa alguna de sus ideas principales con términos elocuentes como el de auto-alienación (2014: “La crisis de la libertad”), y concluye de un análisis correcto que, a diferencia del sometimiento que se da determinado por conceptos de mismidad (la ley) y otredad (la libertad), como en el apartado “La sociedad positiva” (2013), en cambio, la datificación ocurre en los términos de lo equivalente. Sin embargo, sustancializa aquella noción de “poder inteligente”, como una forma nueva de algún dios. Dice: “El poder inteligente lee y evalúa nuestros pensamientos conscientes e inconscientes” (2014, 30). Esta afirmación está excesivamente edulcorada; por un lado, hay individuos detrás de ese “poder” que elaboran las cosas para que otros individuos se datifiquen continuamente (que es la utilidad principal que tiene el “conductismo” de que habla, en tono mesiánico, Lanier). Por otro lado, ese “poder”, en verdad, son los algoritmos que trabajan, en parte, con usuarios. Discutiremos algo más de consideraciones metafísicas semejantes, al final del apartado siguiente que trata sobre el Big Data.

La manera en que se da la orientación que busca la manipulación, puede aproximarse al modo en que algunas empresas generan la “personalización”. Un caso significativo a este respecto es el de Google. Esta empresa es, sin dudas, una base de datos, a los cuales recaba mediante los algoritmos diseñados para rastrear en internet aquello que cada usuario busca. Por cada búsqueda, así como por cada agente que desea que su página esté registrada en Google, este sistema datifica los datos de la red para ordenarlos en pos de ser utilizados en una futura búsqueda (lo cual contribuye a su eficiencia). Al igual que las redes sociales como Facebook, Google “personaliza” las “búsquedas”; dicho de un modo más exacto, va alterando los resultados de las interacciones (que en el buscador de Google se limita exclusivamente a la *apply* consulta) según los datos que recoge de las interacciones anteriores del usuario. A este tipo de ingeniería se le ha dado el nombre más general de “aprendizaje” del sistema (o “*Machine learning*”). Sin embargo,

<sup>13</sup> En el sentido en que Marx había conceptualizado el valor de cambio de la mercancía capitalista en el Tomo I de *El capital* (2008).

<sup>14</sup> Esto debe precavernos de pensar en términos sencillos, como anticipa el título *La jaula del confort* (Magni, 2018), obra en la que el autor analiza variados fenómenos del s. XXI, en una perspectiva entusiasta de la datificación, al sostener que, por ella, se hace una radiografía de la sociedad entera. En este libro, no parece estar correctamente meditado lo que es un dato (si el autor plantea que hay una jaula, hay, entonces, un afuera; pero, como hemos discutido anteriormente, no hay afuera ni adentro, pues no hay distinción entre virtual y real), como tampoco reflexiona en las nuevas características de la sociedad neoliberal (si hay una jaula, entonces, hay una encierro, una limitación; pero, como ya lo hubiera percibido Baudrillard en las *Las estrategias fatales* –2000a–, y últimamente lo haya retomado Han, la contraposición libertad-sujeción, propia de una cultura liberal, ya ha perdido vigencia).

basados en estas consideraciones, podríamos plantear que este aprendizaje no es otra cosa que un continuo reacomodamiento de los resultados de las operaciones algorítmicas según los datos recabados con anterioridad de la red y del usuario. Es una operación de agregación más que de automodificación<sup>15</sup>. De ser así, según parece, el paradigma de aprendizaje de sistema es el mismo que el de la inteligencia artificial: acumulación continua de datos (que implica una continua datificación) y un perfeccionamiento cada vez mayor de su estructura organizacional<sup>16</sup>.

## 4.2 Big Data

Nos interesa explorar otra dimensión de la datificación que cobró gran relevancia en la última década: el Big Data. Este fenómeno propio y exclusivo del proceso de datificación surge como consecuencia suya, pues deriva de la multiplicación enorme de las fuentes de datos *junto* a su fáctica posibilidad de disponerlos en una aglomeración inmensa a través de internet (o en archivos comercializables). Todas las fuentes de datos se dividen entre las que son bases de datos y las que no lo son<sup>17</sup>. La distinción está marcada según el criterio preciso y exacto del

<sup>15</sup> Para ilustrar esta opinión, aquí proponemos un bosquejo del algoritmo de la personalización continua de las búsquedas, en términos del álgebra relacional, para una base de datos como bien podría ser la de búsqueda de Google, o la de los “accesos directos” que expone el sistema operativo Windows (a decir verdad, desconocemos si este es efectivamente la forma en que estas empresas lo crearon, pero, en cualquier caso, es de conjeturar que sean semejantes). Supóngase una base de datos que llamaremos “Rastreadora”. Entre otros posibles esquemas relacionales, para la personalización deberá contener los siguientes (colocaré entre paréntesis sus atributos estructurantes), que son rellenados con la simple utilización de cada usuario con un perfil en ella. La registración del usuario se ubica en la relación USUARIO (‘Nombre de Cuenta’, ‘Nombre’, ‘Apellido’, ‘Fecha de Nacimiento’, ‘Lugar’); el registro de las páginas web en que cada usuario tiene un perfil, en la relación REGISTRACIONES (‘Nombre de cuenta’, ‘Dirección Web’, ‘Temática general de la Página’, ‘Cantidad de Veces Visitada’, ‘Duración de la Visita’); el historial de búsqueda de cada usuario en la relación BUSQUEDA (‘Nombre de cuenta’, ‘Término de Búsqueda’, ‘Dirección Web Seleccionada’, ‘Cantidad de Veces Visitada’, ‘Duración de la Visita’); el registro de páginas web y sus temáticas generales, en PAGINAS (‘Dirección Web’, ‘Temática general de la Página’). Definimos el algoritmo RESULTADO como el conjunto de operaciones algebraicas siguientes: 1. Unión REGISTRACIONES y BUSQUEDA que da un nuevo esquema denominado RG; 2. Concatenación (con las condiciones: ‘Temática general de Página’ = Temática del término de búsqueda actual del usuario, y ‘Término de búsqueda’ = término de búsqueda actual del usuario) RG y PAGINAS, cuyo resultado es un nuevo esquema donde figuran todas las páginas que se asemejen a lo requerido por el usuario, basado en sus búsquedas anteriores. Cuando el usuario decida ingresar a una página del muestrario producido por RESULTADO, ello generará un dato que se adicionará a BUSQUEDA, generando así el proceso de acumulación recursiva.

<sup>16</sup> Esta definición de inteligencia dista mucho de la que nosotros creemos que sea probablemente. Para nosotros, inteligencia es el poder convencer a sí mismo y/o a los otros de algo posible. Jorge Luis Borges es un ejemplo por antonomasia, lo mismo que el Sócrates de los diálogos medios y (algunos) de los tardíos de Platón.

<sup>17</sup> Ejemplos de fuentes de datos son: una central meteorológica, una fábrica integral como Tesla, una fábrica inteligente como Siemens, los comercios de venta en línea, los bancos, los servidores



grado de estructuración del dato en su fuente. Así, siguiendo a Elmasri y Navathe (2007, 784-788), hay datos estructurados, semi-estructurados y no estructurados. Cabe aclarar que no-estructurado no significa nunca que no tiene estructura alguna. La descripción que hicimos de construcción de bases de datos supone, como puede verse, un proceso pormenorizado en la generación de los datos. En cambio, otras fuentes, como algunas páginas web, permiten una producción de datos que no tiene tal ordenamiento, sino el mínimo indispensable para lograr ser útil al fin específico, como en las publicaciones o entradas de un blog. Así, por ejemplo, se ha creado el lenguaje HTML (*HyperText Markup Language*, lenguaje de marcado de hipertexto), porque, partiendo de nociones sencillas, como inicio y finalización de un documento, pueden generarse marcas que están por encima (*hyper*) del contenido, y definen (=delimitan) una estructura mínima.

Cada fuente de datos se caracteriza por el tipo, cualidad y cantidad de datos que produce, tres aspectos que se mantienen aproximadamente invariables dentro de los límites de la actividad humana a la que pertenecen. Las condiciones arriba mencionadas que hacen al Big Data, implican una extralimitación tanto en el tipo como en la cualidad y cantidad de datos<sup>18</sup>, de modo tal que cada uno está ya por fuera de su entorno específico. La extralimitación conlleva que los datos estén disponibles para ser reutilizados, desligados ya de la finalidad particular, limitada, en la que fueron creados. Es importante enfatizar en el hecho de que esta reutilización no parte de que el dato es algo puro en ningún sentido, sino que es necesariamente el resultado de algún tipo de proceso previo<sup>19</sup>. En el Big Data, el dato no ha perdido su identidad, pues, de alguna u otra manera (esto es, según su estructuración, incluso en grado mínimo) remite a su fuente; no obstante, en su condición de extralimitado carece de un sentido e identidad específica. En cierto modo, el Big Data es algo indeterminado, no porque su contenido sea una pura nada, sino porque su “materia” es una aglomeración de datos pre-determinados (por su origen), desfijados, que aún no volvieron a ser fijados a una nueva utilidad, es decir, que todavía no fueron o, mejor dicho, no volvieron a ser procesados. El desafío de la reutilización que supone el Big Data es el de lograr articular la identidad de origen del dato (lo cual implica describirlo correctamente) con una nueva finalidad, una nueva interpretación; en otras palabras, crear un nuevo proceso cuya pretensión principal es hacer lo mismo que los procesos de datificación de los que hemos hablado anteriormente. Por eso,

---

de plataformas, los servidores de hosting, empresas como Google, Baidu, Amazon, Alibaba, Facebook, Telegram, los servidores de internet, las compañías de servicio de comunicación, las páginas web, etc. La autora Pérez Marqués los sintetiza en cinco clases así (2015: “Capítulo 1”): *Web and Social Media, Machine-to-Machine, Big Transaction Data, Biometrics y Human Generated*.

<sup>18</sup> En estas dos se fundamentan dos de las tres V, que la bibliografía reconoce como Volumen y Volatilidad. La tercera, Velocidad, se fundamenta en el ritmo de generación y la consecuente disposición en el espacio virtual, lo cual depende de la conectividad.

<sup>19</sup> Esto fue bien entendido por Srnicek (2018, 42), en su discusión con otros autores respecto del tipo de trabajo que aparece en este “capitalismo de plataformas”, como él lo llama.

en la bibliografía existente sobre Big Data, se proponen lenguajes de interpretación, conformados por algoritmos de análisis y clasificación, que puedan lidiar favorablemente con esa aglomeración indeterminada de datos, entre los cuales se encuentran los lenguajes no-SQL (es decir, no relacionales) u otros como los de tipo entidad/relación, etc<sup>20</sup>.

El resultado continuo e incesante del Big Data, que lo diferencia radicalmente de una base de datos, es la producción de datos *a partir exclusivamente de datos*. El Big Data, pues, puede entenderse como un proceso iterativo de datificación, esto es, una datificación masiva. Es en estas condiciones donde surge una intención metafísica, que expresan algunos entusiastas de la datificación, o que está presente en el término “poder inteligente” de Han, por cuanto se proyecta que la totalidad de las cosas está datificada, es decir, que su entidad está determinada por el dato. Esta pretensión parece contener un equívoco: la creencia de que el Big Data es infinito; pasaremos a explicarlo.

Según hemos visto, en una base de datos, los datos están virtualmente entrelazados con otros datos. Dijimos, también, que esto no podía extenderse de manera infinita, ya que tal entramado está estructurado por limitaciones algorítmicas establecidas durante la creación de la base de datos. Justamente, el Big Data implica la extralimitación de los datos, y esto puede dar lugar a sostener su infinitud. ¿Por qué? Para responderlo y para entender por qué es equivocado, hay que explicitar qué entender por infinito en el contexto de la datificación y extraer sus consecuencias imposibles. Infinito se entiende de dos maneras: o bien, por los resultados posibles de las computaciones informáticas, o por la cantidad en cuanto al almacenamiento. Según el primer sentido, para que el Big Data fuera infinito, se requeriría de operaciones que tengan infinitos resultados (=recursividad infinita); pero esto, de hecho, es imposible, porque se precisa de tales operaciones, de las cuales, justamente, el Big Data en su totalidad carece. Si pudieran encontrarse tales operaciones, se tendría una Gran Base de Datos; pero, no es esa la orientación que tiene este fenómeno informático, sino el de una aglomeración continua y cada vez mayor de datos. A esto último refiere el segundo sentido: cantidad infinita de datos. Para ello se precisaría de una fuente infinita de datos. Esta es una de las pretensiones o idealidades prácticas de la datificación, por cuanto aspira a que cada cosa que hay en el mundo y cuyo movimiento pueda ser consignado en etapas, envíe una señal que sea receptada y luego datificada

---

<sup>20</sup> Para un acercamiento al problema, ver Tabares, L. F. y Hernández, J. F. (s/a); también Sousa, J. S., Miranda Martins, P. G. y Sá Ramalho, R. A. (2018). ¿Por qué se suele hacer tanto énfasis en lenguajes no-SQL? No solo esto indica la importancia y el uso extendido de las bases relacionales, sino que, además, señala el carácter de aglomeración de los datos. En el Big Data se disponen de datos de cualidades múltiples (como instancias, tuplas, columnas, etc) y de grados diversos de estructuración, que, tomados en su totalidad (o incluso en una porción de éstos), resultan definitivamente divergentes. Por lo tanto, no es posible aplicar, en primera instancia, un análisis relacional, ya que, para que los esquemas relacionales sean *efectivamente útiles*, precisan de cierta coherencia entre sus atributos, y es esa coherencia la que hay que hallar primero en el Big Data.

(la Internet de las Cosas). Pero esto es físicamente imposible, ya que se precisa tanto de una multiplicidad tecnológica que recepcione cada una de tales cosas, así como de la instalación de sensores por doquier. Por lo tanto, tampoco puede ser infinito en cantidad.

El Big Data refleja, pues, intenciones metafísicas, totalizantes, e imposibles. No obstante, esta conclusión merece una salvedad. Porque bien puede ser prácticamente imposible, y que probablemente permanezca en una forma nebulosa (difusa), que resulte parcial y momentáneamente utilizable. Sin embargo, creemos que es un fenómeno que tiende a hegemonizar la forma de darse las actividades humanas, que podría derivar en una aparente consumación de su instalación total<sup>21</sup>.

## 5. Conclusiones

A lo largo de este artículo hemos presentado y desarrollado algunos de los conceptos más relevantes de la ciencia de la informática, para establecer discusiones filosóficas respecto de lo que es un dato, de cómo sea la relación dato-mundo, de la transformación que conlleva la datificación en la subjetividad y las relaciones sociales, como son las redes sociales, y, finalmente, cómo entender la masificación de los datos a través del Big Data. Hemos también revisado críticamente algunas de las consideraciones hallables en la bibliografía sobre el asunto, como son la idea de que el dato es un reflejo del mundo, o que la datificación re-edita aparatos de control que eran propios de las sociedades liberales. Creemos que algunos de estos lineamientos permitirán seguir discutiendo, no solo los mismos temas que aún merecen sin duda mayor reflexión, sino otros que son igualmente relevantes y que no hemos podido abordar por las limitaciones propias de un artículo, como es la relación entre la datificación y la ciencia, u otras repercusiones de la metafísica moderna (como es la relación sujeto-objeto). Lo cual será materia de posteriores trabajos.

---

<sup>21</sup> Además de esta idealización interna del desenvolvimiento de la datificación, la hegemonización deriva, también, por otros factores, como las grandes ganancias que representan para algunas empresas especialmente dedicadas a explotar la datificación (como Alibaba o Amazon) y las empresas productoras y/o comerciales de las tecnologías para ello (como Samsung o Huawei), de las facilidades organizativas que supone en algunas industrias (como Tesla o Siemens), de la importancia que adquiere la datificación para un mundo definitivamente capitalista y para la hegemonía internacional. Para un informe de algunos de estos aspectos, remitimos al documento citado de la CEPAL. Para un acercamiento a las implicancias en política internacional, Zuccaro, A. y Schulz, S. (2020) escribieron un estado de la cuestión interesante, aunque contengan apreciaciones sobre la datificación equivocadas.

## Referencias bibliográficas

- Anderson, C. (2008). *El fin de las teorías: el diluvio de datos hace obsoleto al método científico*. <https://simonjosepulido.wordpress.com/2016/04/10/el-fin-de-las-teorias-el-diluvio-de-datos-hace-obsoleto-al-metodo-cientifico/>
- Aristóteles. (2009). *Física*. Barcelona: Gredos.
- Baudrillard, J. (2000a). *Las estrategias fatales*. Barcelona: Anagrama.
- Baudrillard, J. (2000b). *El crimen perfecto*. Barcelona: Anagrama.
- Castells, M. (2000). *La era de la información: economía, sociedad y cultura*. Madrid: Alianza.
- CEPAL. (2015). *La nueva revolución digital. De la Internet del consumo a la internet de la producción*. Santiago: CEPAL.
- Codd, E. F. (1970). A relational model of Data for Large Shared Data Banks. *Communications of the ACM*, 13(13), 377-387.
- Diéguez, A. (2017). *Transhumanismo*. Barcelona: Herder.
- Elmasri, R., Navathe, S. B. (2007). *Fundamentos de sistemas de bases de datos*. Madrid: Pearson Educación.
- Frege, G. (1972). *Conceptografía*. Ciudad de México: Universidad Nacional Autónoma de México.
- Frege, G. (1984). *Estudios sobre semántica*. Barcelona: Orbis.
- Gamut, L. T. F. (2002). *Introducción a la lógica*. Buenos Aires: Eudeba.
- Granroth, J. (2020). *Technological imaginaries and technological determinism: algorithms mediating a better tomorrow?* Tesis. Facultad de Ciencias Sociales, Universidad de Helsinki.
- Han, B.-C. (2013). *La sociedad de la transparencia*. Barcelona: Herder.
- Han, B.-C. (2014). *Psicopolítica*. Barcelona: Herder.
- Lanier, J. (2018). *Diez razones para borrar tus redes sociales de inmediato*. Madrid: Debate.
- Magni, E. (2019). *La jaula de confort*. Buenos Aires: Autoría Sherpa.
- Marx, K. (2008). *El capital, Tomo I, Vol. I*. Buenos Aires: Siglo XXI.
- Mayer-Schönberger, V., Cukier, K. (2013). *Big Data. La revolución de los datos masivos*. Madrid: Turner.
- Morozov, E. (2011). *El desengaño de internet, los mitos de la libertad en la red*. Barcelona: Destino.

- Pérez Marqués, M. (2015). *Big data. Técnicas, herramientas y aplicaciones*. Santiago: Alfaomega.
- Srnicek, N. (2018). *Capitalismo de plataformas*. Buenos Aires: Caja Negra.
- Sousa, J. L., Martins, P. G. M., Ramalho, R. A. Sá (2018). Modelos de representação semântica na era do Big Data. *Brazilian Journal of Information Studies: Research Trends*, (12), 34-40.
- Tabares, L. F., Hernández, F. J. (s/a). *Big Data analytics: Oportunidades, retos y tendencias*. [https://www.academia.edu/15043636/Big\\_Data\\_Analytics\\_Oportunidades\\_Retos\\_y\\_Tendencias](https://www.academia.edu/15043636/Big_Data_Analytics_Oportunidades_Retos_y_Tendencias)
- Valdés Villanueva, L. M. (Comp). (2005). *La búsqueda del significado*. Madrid: Tecnos.
- van Dijk, J. (2014). Datafication, Dataism and Dataveillance: Big Data between Scientific Paradigm and Ideology. *Surveillance and Society*, 12(2), 197-208.
- Zuccaro, A., Schulz, S. (2020). La disputa por la hegemonía mundial: *big data*, cuarta revolución industrial y 5G. *Actas de Periodismo y Comunicación*, 6(2).