



# El imperio

Xavier Duran

# de los datos

El Big Data, la privacidad  
y la sociedad del futuro

# El imperio de los datos

El Big Data, la privacidad  
y la sociedad del futuro

Xavier Duran

PREMIO EUROPEO DE DIVULGACIÓN CIENTÍFICA  
ESTUDI GENERAL 2017



Directora de la colección:  
Carolina Moreno

Coordinación:  
Soledad Rubio

*Esta publicación no puede ser reproducida, ni total ni parcialmente, ni registrada en, o transmitida por, un sistema de recuperación de información, en ninguna forma ni por ningún medio, ya sea fotomecánico, fotoquímico, electrónico, por fotocopia o por cualquier otro, sin el permiso previo de la editorial. Diríjase a CEDRO (Centro Español de Derechos Reprográficos, [www.cedro.org](http://www.cedro.org)) si necesita fotocopiar o escanear algún fragmento de esta obra.*

© Del texto: Xavier Duran Escriba, 2019

© De la presente edición:

Unitat de Cultura Científica  
i de la Innovació de la Universitat de València  
[www.valencia.edu/cdciencia](http://www.valencia.edu/cdciencia)  
[cdciencia@uv.es](mailto:cdciencia@uv.es)

Publicacions de la Universitat de València, 2019

[www.uv.es/publicacions](http://www.uv.es/publicacions)  
[publicacions@uv.es](mailto:publicacions@uv.es)

Producción editorial: Maite Simón

*Interior*

Diseño y maquetación: Inmaculada Mesa  
Corrección: Letras y Píxeles S.L.

*Cubierta*

Diseño original: Enric Solbes  
Grafismo: Celso Hernández de la Figuera

ISBN: 978-84-9134-480-3

La tecnología no es ni buena, ni mala, ni neutral.

A pesar de que la tecnología puede ser un elemento principal en muchos asuntos públicos, los factores no técnicos deben tener preferencia en las decisiones sobre política tecnológica.

(1.ª y 4.ª leyes de Kranzberg)  
MELVIN KRANZBERG (1986)

El objetivo principal de toda ciencia es la libertad y la felicidad del hombre.

THOMAS JEFFERSON (1810)

Es un gran error elaborar teorías antes de tener datos. Inconscientemente, empiezas a distorsionar los hechos para adaptarlos a la teoría en vez de adaptar la teoría a los hechos.

(Sherlock Holmes)  
ARTHUR CONAN DOYLE  
(*Escándalo en Bohemia*)

## *Premios Literarios Ciutat d'Alzira 2017*

Esta obra obtuvo el XXIII Premio Europeo de Divulgación Científica Estudi General, instituido por la Universitat de València y el Ayuntamiento de Alzira. Formaban parte del jurado Carmen Agustí, Pilar Campins, Andreu Escrivà, Lucía Hipólito y Fernando Sapiña.

# ÍNDICE

## **INTRODUCCIÓN: YO SOY YO Y MIS DATOS**

### **Capítulo 1. VIAJE AL PAÍS DE LOS DATOS**

EL NACIMIENTO DE LOS DATOS

UNIDADES DE INFORMACIÓN Y SUS EQUIVALENCIAS

TODA LA INFORMACIÓN EN UNOS Y CEROS

LAS CINCO V DEL BIG DATA

CÓMO PROCESAR LOS DATOS

### **Capítulo 2. TODO LO QUE NOS CUENTA EL BIG DATA**

DEL COSMOS AL ÁTOMO

CEREBRO, CÁNCER, CLIMA, QUÍMICA

LOS ALGORITMOS TRABAJAN

CONTROLANDO MILLONES DE PACIENTES A LA VEZ

### **Capítulo 3. UN MUNDO (MÁS O MENOS) FELIZ**

LA CIUDAD INTELIGENTE

CON LOS DATOS PUESTOS

EL DOCTOR GOOGLE TE VISITA

INTERLUDIO DE FICCIÓN: UNA MAÑANA CUALQUIERA DEL SEÑOR PUIG

### **Capítulo 4. SABEN LO QUE HAS HECHO Y LO QUE HARÁS**

SABEN MÁS DE TI QUE TÚ MISMO

HOGAR, DIGITALIZADO HOGAR

EL PRECIO DE LOS DATOS

LAS REDES TE CONOCEN MEJOR QUE LA PAREJA

## **Capítulo 5. CIUDADANOS CLASIFICADOS**

LOS CENSOS SE MODERNIZAN: DEL LÁPIZ AL SATÉLITE

LAS DUDAS VIENEN DE LEJOS

PONER A LOS CIUDADANOS EN CUBETAS

SEGUROS POCO SEGUROS

¿QUÉ DICEN LOS GENES?

¿CUÁNTOS AÑOS VIVIRÁ ESTA PERSONA?

## **Capítulo 6. REDES CONTRA EL DELITO**

¿ALGORITMOS RACISTAS?

SU CARA LE SUENA A MI ALGORITMO

IDENTIFICARSE POR LA CARA

INTERLUDIO DE FICCIÓN: LECCIONES DE INGENIERÍA DOMÉSTICA

## **Capítulo 7. MEGADATOS Y MEGAERRORES**

EL DEMONIO DE LOS NÚMEROS

LOS DATOS NO PIENSAN

HUMANOS Y MÁQUINAS

## **Capítulo 8. HISTORIAS DEL LADO OSCURO**

PIRATAS DE LA WEB

CIBERBARRERAS CONTRA EL CIBERDELITO

INTERLUDIO DE FICCIÓN: MUERTE FÍSICA, VIDA DIGITAL

## **Capítulo 9. LA HORA DE LOS DERECHOS DE LOS USUARIOS**

ANÓNIMOS, PERO NO TANTO

DATOS PARA LA ETERNIDAD

UN OCÉANO DE DERECHOS

## **Capítulo 10. CONCLUSIONES EN FORMA DE DECÁLOGO**

BIBLIOGRAFÍA

ÍNDICE ANALÍTICO

## Introducción

### YO SOY YO Y MIS DATOS

Datos, datos, datos a montones... Vivimos en un mundo de datos, almacenados en formas variadas: textos, números, imágenes, gráficos... «Yo soy yo, mis circunstancias... y mis datos», diría hoy Ortega y Gasset.

Hay tantos y tantos que ya no hablamos de datos, sino de Big Data, grandes datos. El concepto ha hecho fortuna y pese a que a menudo se deja en inglés, también se adapta a cada idioma. En castellano se habla de megadatos o de datos masivos. Utilizaremos preferentemente Big Data, pero también usaremos las dos traducciones. La idea siempre es que hay muchos datos.

Big Data nos hace pensar en archivos digitales y en consultas por internet. Pensamos en Google y en las montañas de información por donde debe moverse este buscador para buscar lo que le pedimos. Y quizá pensamos en Facebook y en Instagram. Pero, como iremos explicando a lo largo del libro, parece que nada queda al margen del Big Data: ni mensajes privados por WhatsApp, ni llamadas telefónicas, ni compras con tarjeta, ni siquiera los paseos con el móvil encendido. Pero hay muchas más fuentes de datos: los que mandan los satélites, los que proporcionan sensores repartidos por las ciudades, por el campo o por los océanos, las imágenes de cámaras de seguridad, los datos que proporcionan aparatos médicos o los llamados *wearables* -una especie de captadores de datos portátiles, que pueden consistir en un brazalete o en una prenda, como una camiseta.

Ya escribió el filósofo inglés Francis Bacon, a finales del siglo XVI, que «Conocimientos es poder». Pero datos y

conocimiento no son lo mismo. De hecho, incluso hay entre ellos un paso intermedio, que es la información. Confundimos datos con información y son cosas distintas. Un grupo de músicos tocando por su cuenta, por afinadamente que lo hagan y por virtuosos que sean, son datos. Todos ellos tocando en armonía a las órdenes de un director de orquesta es información.

Los datos son el combustible que permite resolver problemas -a veces, creados por los datos mismos-. Pero un combustible solo no sirve de nada. Los datos sirven para que funcione la maquinaria que busca las respuestas a los problemas. Por eso, los datos son imprescindibles, pero sin una estrategia para tratarlos y transformarlos no tendríamos nunca información. Y una vez reunida suficiente información, aún nos queda el trabajo de analizarla y de reflexionar. De la manera como la utilizemos para producir conocimiento dependerá la calidad de este.

Aun así, no podemos negar que, hoy en día, los datos son poder. Hay quien los llama «el petróleo del siglo XXI». Volvemos a la metáfora del combustible, pero en este caso para alimentar máquinas de fabricar dinero -y de construir poder-. Quien tiene muchos datos tiene mucho poder, si sabe cómo utilizarlos... o si los vende a alguien a quien le interese hacerlo.

Aquí explicaremos de dónde surgen tantos datos, cómo circulan, cómo se guardan. Mostraremos cómo se procesan -algo que se puede hacer bien o muy mal-. Y describiremos los beneficios que aportan y los riesgos que representan. Muchos posibles beneficios y muchos posibles riesgos. Algunos ya son palpables -tanto las derivaciones positivas como los peligros- y otros están a punto de llegar, aunque parezcan fantasías de película de serie B.

En definitiva, proporcionaremos al lector muchos datos, transformados en información, con la esperanza de que generen conocimiento. No sabemos si nuestra aportación será valiosa, pero no tenemos ninguna duda de que

intentarlo es necesario. Pueden existir datos sin información, pero difícilmente habrá información sin datos. Y aún menos, conocimiento. Para que el imperio de los datos no nos engulla, hay que estar medianamente preparados. Solamente si los ciudadanos tienen suficientes datos y los saben procesar podrán presionar para que la información y el conocimiento que se derivan de ellos sean beneficiosos para la sociedad.

# Capítulo 1

## VIAJE AL PAÍS DE LOS DATOS

Había 5 exabytes de información creados desde el alba de la civilización hasta 2003, pero esta información ahora se genera cada dos días.

ERIC SCHMIDT (2010)

El mundo ya no está dominado por las armas, ni por la energía, ni por el dinero. Está dominado por unos y ceros, por pequeños bits de datos. Todo está en los electrones.

COSMO, personaje de la película  
*The sneakers (Los fisgones, 1992)*

A lo largo del siglo XX han tenido gran repercusión tres conceptos científicos profundamente desestabilizadores que lo han dividido en tres partes desiguales: el átomo, el bit y el gen. [...] Cada uno tiene su origen en una noción científica abstracta, pero crece hasta acabar invadiendo un gran número de disciplinas humanas y transformando la cultura, la sociedad, la política y el lenguaje.

SIDDHARTHA MUKHERJEE

Fremont Rider levantó la vista para contemplar las estanterías llenas de libros, suspiró e inmediatamente pensó en un futuro más bien negro o, por lo menos, muy complejo. Rider era escritor y bibliotecario de la Universidad Wesleyana en Middleton (Connecticut, Estados Unidos) y en el año 1944 lanzó un grito de alarma respecto a la cantidad de libros que se publicaban anualmente. Calculó que las bibliotecas norteamericanas duplicaban su tamaño cada dieciséis años. Según Rider, a este ritmo, la biblioteca de la Universidad de Yale, una de las principales del país, tendría, en el año 2040, «aproximadamente 200.000.000 de volúmenes, que ocuparían 9.656 kilómetros de estanterías». El problema no sería solo de espacio, sino también de gestión. Rider calculaba que esta cantidad de libros haría necesario un equipo de más de seis mil personas para catalogarlos.

Más de siete décadas después del aviso de Rider, el problema ya no son tanto los libros editados como el conjunto de la información. Internet ha provocado una explosión de datos. Solamente con los que procesa cada día Google se podrían editar volúmenes suficientes para que, apilados, llegasen a la mitad de camino entre la Tierra y la Luna. Quizá Rider ni tan solo tendría ánimos de calcular cuánto personal se necesitaría para catalogarlos -una sencilla regla de tres con los datos del bibliotecario americano revela que serían más de 118.000 personas.

Afortunadamente, estos datos no se encuentran en papel, sino que más del 90 % se hallan en soporte digital. Desgraciadamente, no tenemos que considerar solo las búsquedas en Google, sino todo lo que se genera en el universo digital en distintos formatos.

De vez en cuando, alguien realiza cálculos parecidos a los de Rider, pero ya no se pueden limitar al papel. Además, suelen quedar obsoletos al cabo de poco tiempo. En 1997, Michael Lesk, un informático y experto en sistemas de información, se entretuvo en calcular cuánta información

existía en el mundo. Empezó describiendo la Biblioteca del Congreso en Washington, con sus veinte millones de libros, trece millones de fotografías, cuatro millones de mapas, más de medio millón de películas y tres millones y medio de registros de sonido.

Pero Lesk no se podía limitar a una biblioteca, por grande que fuera, ni tan solo al material editado. Añadía que en un año se filmaban miles de películas, se realizaban miles de millones de fotografías, se emitían millones de horas de televisión y de radio, se editaban más de 400 millones de CD y más de 300 millones de casetes -muchos duplicados, sin duda, porque de algunos se hacían miles de copias-, había billones de minutos de conversaciones telefónicas... Realizando cálculos aproximados y basándose en otras fuentes, señalaba que quizá en el mundo había 12.000 petabytes (PB) de información. Esto significa 12.000 millones de gigas, por usar una unidad de medida que a mucha gente le resulta familiar.

Pese a estas cifras, concluía que en la Tierra habría suficiente capacidad de almacenamiento para todo lo que la gente escribiese, dijese, fotografiase o filmase en el futuro.

De todo ello se cumplen algo más de veinte años y la cantidad de información ha aumentado de forma exponencial. Y parece que sí, que la tecnología, al menos de momento, está solucionando el problema de guardarla e incluso de hacerla accesible. Pero ¿qué utilidad puede tener tanta información? ¿Y cómo podemos gestionarla?

## EL NACIMIENTO DE LOS DATOS

Los datos nacen de la necesidad. Hubo datos antes de que hubiese métodos para representarlos de forma comprensible para todo el mundo. Primero fueron los datos y, tiempo después, aparecieron los números. Hace miles de

años, un pastor veía que de su corral salían muchas ovejas y que después de pasturar entraban muchas. Pero ¿cómo podía saber si volvían todas?

Para estar seguro de que no perdía ninguna oveja debía tomar una piedra o una ramita por cada una que salía del corral. Y cuando después de pasturar volvían a entrar, debía retirar una piedra o ramita del montón por cada una. Si no quedaba ninguna, todas habían vuelto. Si quedaban piedras en el montón, alguna se había escapado. Y si seguían llegando ovejas y ya había acabado las piedras y las ramitas, o bien se había descontado, o bien había ganado algún ejemplar extra.

Más tarde llegarían los sistemas para simbolizar las cantidades. Las sociedades evolucionaban, se hacían más complejas. Había más producción agrícola y había más rebaños. Y se hacían intercambios comerciales. Así nacieron los números. No los números actuales, sino otros sistemas simbólicos para representar cantidades. Hace más de cinco mil años ya había fichas de arcilla con símbolos que correspondían a cantidades e incluso a cálculos.

Pero la información, los datos, no era simplemente numérica. Había textos, había representaciones simbólicas, había grabados. Ahorrémonos unos cuantos milenios y saltemos al siglo xv. Con la imprenta, la información editada con libros y documentos estalla y hay quien ve un alud difícil de gestionar. Los primeros escépticos sobre la capacidad humana para asimilar tantos libros no pudieron ver que cualquiera de sus previsiones se quedaba corta en pocas décadas.

Hagamos nuevamente un gran salto. A mediados del siglo xx, la cantidad de información era inmensa y a alguien se le ocurrió que tenía que haber alguna manera de cuantificarla. En 1948, el norteamericano John W. Tukey, matemático y pionero de la informática, creó el bit, como abreviatura de *Binary digiT*. Aparte de la contracción del concepto en tres letras, debía jugar con el significado de bit

en inglés, 'pieza pequeña'. Ya tenemos la unidad de información digital.

Al cabo de pocos años, en 1956, el ingeniero electrónico Werner Buchholz -norteamericano nacido en Alemania, de donde se marchó huyendo del nazismo- creó el *byte*. En los años cincuenta, Buchholz trabajaba en la IBM y formó parte del equipo que diseñó los primeros ordenadores, como el IBM 701. El bit era demasiado pequeño para medir la cantidad mínima de información, un solo carácter, y por eso surgió el byte. Al principio, no había una equivalencia estándar y un byte, según el sistema o el ordenador utilizados, podía variar. Ahora, un byte equivale a 8 bits y por eso a veces se le llama octeto.

Ya tenemos el byte, pero pese a la necesidad de definir la unidad que equivale a un solo carácter, una medida tan pequeña tiene poca utilidad cuando hablamos de grandes cantidades de información. Sería como medir distancias astronómicas en centímetros. Por ello, en seguida surgieron los múltiplos: kilobyte, megabyte... Pero mega, un millón, se queda corto en muchos casos y por eso aparecieron el giga (mil millones) y otros que progresivamente multiplican el anterior por mil: tera, peta, exa, zetta, yotta... Con este último llegamos al cuatrillón.

Explicábamos antes que Lesk había situado en 12.000 petabytes la cantidad de información que había en el mundo en 1997. Pero con estas cifras a mucha gente le pasa como con los presupuestos estatales o con los beneficios de las grandes empresas. Nos pueden hablar de 17.000 millones de euros, de 80.000 millones o de 250.000 millones. Comprendemos que es muchísimo, pero somos incapaces de hacernos una idea.

Por eso, algunas comparaciones serán útiles. Un byte es un solo carácter. Por tanto, una sola letra ocupa un byte. Si creamos un documento con una sola letra, «pesará» un byte. A partir de aquí, el primer paso no es difícil. Un

kilobyte (KB) equivale a media página, unos mil caracteres. Y un megabyte podría ser una novela corta.

Hagamos un breve inciso. A menudo leemos que 1 KB son 1.024 bytes. Esto se debe al origen del byte y a que los informáticos trabajan en sistema binario y, por lo tanto, con potencias de dos. Como  $2^{10}$  es 1.024, esta es la equivalencia que se utiliza a menudo en el ámbito de los ordenadores. Pero para el sistema internacional de medidas, 1 KB son mil bytes.

Pero la información no está solo en forma de texto o de cifras. Podemos tener gráficos, dibujos, fotografías... Incluso películas o sonidos. Cada añadido aumenta la cantidad de información. Una fotografía con buena definición puede ocupar dos megabytes. Es decir, como dos novelas cortas.

Una hilera de diez metros de libros equivale a un gigabyte (GB). Y con seis millones de libros tendríamos un terabyte (TB). Si reuniéramos siete millones de horas de televisión de alta definición tendríamos un petabyte (PB). ¡Y Lesk decía que toda la información que había en el mundo ocupaba 12.000 petabytes! Hoy, en tan solo una hora ya se transmiten en todo el mundo 500 petabytes de información, equivalentes a 6.600 años de vídeo de alta definición o a diez veces todas las obras escritas por la humanidad desde los inicios de la historia.

Todas estas comparaciones son aproximadas. La cantidad de bytes que tiene un texto también depende de las órdenes de estilo que incorpore -formato, estilo y tamaño de letra...-. Una fotografía puede tener mucha calidad o muy poca y lo mismo pasa con una película. Por otro lado, se hacen comparaciones con cosas muy difíciles de medir con exactitud. Así, se ha dicho que todas las palabras pronunciadas por toda la humanidad a lo largo de la historia ocuparían cinco exabytes (EB). La idea también ha sido rebatida y nuevos cálculos hablan de 42 zetabytes (ZB). Pero es muy probable que nos falten muchos elementos para poder valorarlo con precisión.

# UNIDADES DE INFORMACIÓN Y SUS EQUIVALENCIAS

(Cada una multiplica por mil la anterior)

- 1 byte 1 carácter
- 1 kilobyte Media página mecanografiada  
(KB)
- 1 megabyte Una novela corta  
(MB)
- 1 gigabyte Una película de dos horas  
(GB)
- 1 terabyte Seis millones de libros  
(TB)
- 1 petabyte 2.000 años seguidos de música  
(PB)
- 1 exabyte 100.000 veces todo el material impreso -libros,  
(EB) revistas, documentos- de la Biblioteca del Congreso de  
Washington
- 1 zetabyte 152 millones de años de vídeo de alta definición  
(ZB)
- 1 yotabyte Toda la información que puede contener el centro de  
(YB) datos de la NSA (National Security Agency) de Estados  
Unidos en Utah, que tiene una superficie de 92.000  
metros cuadrados

## TODA LA INFORMACIÓN EN UNOS Y CEROS

Algo que sí se puede calcular con más certeza, aunque también tendrá imprecisiones, es la capacidad de almacenaje de la información. Así, en 1986 se podían guardar, con los dispositivos existentes en todo el mundo, 2,6 exabytes, y en 2007 ya podían ser 295 EB. Esto significa que en 1986 había el equivalente a menos de un CD por persona y en 2007 ya eran unos 61 CD por persona (Marinescu, 2013: 196) -no es una relación lineal porque la capacidad había aumentado, pero la población del planeta también-. En total se podrían haber llenado más de 400.000

millones de CD, que apilados ocuparían una distancia superior a la que hay entre la Tierra y la Luna.

Si la cantidad de información aumenta, también debe hacerlo la capacidad para almacenarla. En 2009, el ingeniero norteamericano Mark Kryder enunció la ley que lleva su nombre. Pronosticaba que si los discos duros continuaban progresando al mismo ritmo que en los años anteriores, aumentando la capacidad un 40 % anual, en el año 2020 dos discos de 2,5 pulgadas podrían contener unos 40 terabytes y valdrían unos 40 dólares -unos 36 euros actuales-. No parece que la ley se cumpla, porque la capacidad de almacenaje tan solo se había doblado en 2014, cinco años después de que se enunciase la ley. El aumento no es lento, pero es menor que el ritmo de crecimiento de la información generada.

Hay que tener en cuenta, sin embargo, que siempre pueden aparecer nuevas técnicas. La más innovadora es la que utiliza el ADN para guardar información y reproducirla. Dicho de manera sencilla y simplificando, primero se trata de digitalizar las imágenes -fotografías o películas- que queramos guardar. A cada píxel se le hace corresponder una secuencia concreta de ADN, según su color -blanco, negro o todas las tonalidades de gris-. Recordemos que las cadenas de ADN están formadas por cuatro bases -adenosina, citosina, guanina y timina, simbolizadas por las letras A, C, G y T-. Así, por poner un ejemplo, un cierto tono de gris se haría corresponder con el fragmento ATC y otro tono con GTC. Habríamos transformado la imagen que estaba en código digital en un código de ADN. Cada secuencia lleva, además, una etiqueta que permite saber cómo se tiene que colocar cuando se reproduzca la fotografía.

Así, obtenemos una especie de mapa de ADN que nos indica qué tono tiene cada uno de los píxeles de la imagen y dónde está situado. Después se sintetiza la secuencia de ADN que hemos establecido. Cuando alguien quiera recuperar la información -la imagen-, lo podrá hacer a partir

de este ADN. Deberá secuenciar la cadena y hacer corresponder cada fragmento con una tonalidad.

La información que se guarda con este sistema no tiene por qué estar en forma de imágenes. Podemos guardar textos o archivos. La clave es simplemente establecer la equivalencia entre un píxel, una palabra o un conjunto de palabras con una secuencia de ADN. En 2018 se publicaron dos artículos que describían experiencias reales. El primero explicaba cómo se almacenaron en ADN seis archivos, entre ellos un sistema operativo completo y una de las primeras películas de cine que se hicieron -*Llegada del tren a la estación de La Ciotat*, de los hermanos Lumière (Erich y Zielinski, 2017)-. En total se utilizaron 72.000 fragmentos de ADN, cada uno con una longitud de 200 bases. Según los autores, su método permitía guardar 215 petabytes en un gramo de ADN.

En el segundo estudio se almacenó en ADN la secuencia de imágenes obtenida por el fotógrafo británico Eadweard Muybridge en 1878, que tenía como objetivo averiguar si en algún momento las cuatro patas de un caballo al galope estaban todas en el aire. En este caso, no solo se guardaron las imágenes en secuencias de ADN, sino que después se introdujeron en el genoma de la bacteria *E. coli* (Shipman *et al.*, 2017). Los investigadores comprobaron que las generaciones sucesivas de bacterias conservaban esta información y que la podían volver a traducir para recuperar las imágenes.

Esto elimina uno de los grandes problemas del almacenaje de información: que los soportes donde se guarda queden obsoletos y no existan los aparatos para volverla a leer al cabo de unos años. Si la información se guarda en forma de ADN en bacterias, siempre se podrá secuenciar el ADN y recuperarla -si se conoce el código de equivalencia, claro.

En teoría, el ADN permitiría guardar en una sola habitación toda la información producida por la humanidad

a lo largo de su historia (Service, 2017). De momento, aparte de los problemas técnicos que se deben ir superando, la dificultad consiste en el coste. Sintetizar ADN que contenga dos megabytes cuesta 7.000 dólares y secuenciarlo para recuperar la información 2.000 más. Además, el sistema es más lento que leer directamente la información que hay en un CD o un lápiz de memoria. Pero aquí lo exponemos como un nuevo camino que se abre para guardar información.

Al mismo tiempo que va aumentando la capacidad de almacenar información, lo hace, de manera extraordinaria, la de transmitirla y recibirla. Según la ley de Nielsen, enunciada en 1998 por el danés Jakob Nielsen, la capacidad de las redes se dobla cada 21 meses. Todas estas previsiones, como la conocida ley de Moore, según la cual la densidad de los chips de silicio se dobla cada 18 meses, son aproximaciones que tal vez no se cumplan siempre o algún día dejen de hacerlo. Por lo tanto, pueden ser orientativas en cierto momento, pero no se pueden tomar como previsiones de futuro a medio o largo plazo. Pese a ello, inexactitudes al margen, hay algo que parece claro: producimos más información que nunca y, además, tenemos capacidad para hacerla viajar de un lado a otro por todo el planeta.

Y esto último es posible porque existe otra diferencia respecto a épocas anteriores: la mayor parte de la información está digitalizada. No tenemos solamente libros, revistas, fotografías, películas, sonidos... Tenemos todo esto en un formato que permite guardarlo en ordenadores y en dispositivos de memoria e intercambiarla con facilidad. Ya no es necesario -aunque todavía lo hacemos- pedir un disco y esperar a que nos llegue por correo para escucharlo. Instantáneamente podemos buscar la música o lo que sea y bajarla a nuestro ordenador. Cuestión de minutos, como mucho.

Este es el gran cambio respecto al mundo analógico. Ya no se trata de revolver libros o archivos fotográficos -que también hay que hacerlo y se sigue haciendo-. Ya no removemos montones de manuscritos o carretes de microfilms. La mayor parte de datos ya se crean en formato digital. Y los que se hallan en analógico se van digitalizando. Es esto lo que hace posible hablar de Big Data y utilizar estos datos.

Por lo tanto, allí donde vemos datos, en realidad solo hay ceros y unos, montones de información traspasados al sistema binario. Evidentemente, se trata de unos montones de ceros y unos muy complejos, con grandes matrices multidimensionales y vectores que permiten tener una fotografía o una película entera guardada en un dispositivo y pasarla de un lugar a otro.

¿Pero dónde están todos estos datos? Algunos los guardamos en los discos duros o en dispositivos como lápices de memoria. Una gran parte los encontramos cuando navegamos por la red. A veces oímos decir que están en la nube -*the cloud*-. No significa que los datos -o incluso programas informáticos- viajen por el espacio electromagnético hasta que alguien los recupera en su ordenador, móvil o tableta. ¿Qué significa nube?

Pese a que los datos no sean más que complejos conjuntos de ceros y unos, necesitan un soporte físico. Por lo tanto, los datos se hallan en sitios concretos. Son los grandes servidores, los grandes centros de datos. Tienen servidores las grandes empresas y los tienen las no tan grandes. Y para gigantes como Amazon, Google o Facebook hacen falta grandes edificios, llenos de procesadores y de cables. Y como aparte de ser grandes tienen mucha visión comercial y mucho poder, también alquilan espacio a empresas o instituciones que no tienen sus «granjas de datos» porque no pueden -o no quieren- asumir los costes que representa construir unos edificios con mucho espacio y grandes medidas de seguridad -allí se encuentran sus