



## ARTÍCULO ESPECIAL

### *Big data*, ¿pero qué es?

R.M. Moreno-Carriles



Servicio de Angiología y Cirugía Vasculard, Hospital Universitario La Princesa, Madrid, España

Recibido el 20 de marzo de 2018; aceptado el 2 de mayo de 2018

Disponible en Internet el 7 de junio de 2018

#### PALABRAS CLAVE

*Big data*;  
Análisis sistemático  
computarizado de  
datos;  
Atención sanitaria;  
Cirugía vascular

**Resumen** El *big data* puede aplicarse a campos tan diversos como la seguridad, la administración pública, la logística o la relación con el cliente. Puede revolucionar el transporte inteligente, el ahorro energético, la conservación del medio ambiente, los sistemas de análisis financiero, la educación, la ciencia, la investigación y el cuidado de la salud. Es un nuevo paradigma en el que los cirujanos vasculares debemos integrarnos.

© 2018 SEACV. Publicado por Elsevier España, S.L.U. Todos los derechos reservados.

#### KEYWORDS

*Big data*;  
Analytics;  
Health care;  
Vascular surgery

#### **Big data; But what is it?**

**Abstract** *Big data* can be applied to fields as diverse as security, public administration, logistics, or customer relations. It can revolutionise intelligent transport, energy saving, environmental conservation, financial analysis systems, education, science, research and health care. It is a new paradigm into which vascular surgeons must integrate.

© 2018 SEACV. Published by Elsevier España, S.L.U. All rights reserved.

#### Definición

Macrodatos e inteligencia de datos son alternativas en español a la voz inglesa *big data*, que se emplea en el sector de las tecnologías de la información y de la comunicación (TIC) para aludir a un conjunto de datos que, por su volumen, variedad y velocidad de producción, no pueden ser analizados utilizando procesos o herramientas tradicionales.

El término macrodatos es una alternativa válida, aporta, como *big*, el significado de «grande»; es una solución breve y no tiene, como ocurriría con megadatos, un posible riesgo de confusión con el término mega, muy frecuente también en los mismos ámbitos.

En cuanto a inteligencia de datos, se trata de otra opción igualmente válida, centrada en el concepto que subyace al mundo del *big data*, donde las ingentes cantidades de información que se manejan pueden generar «inteligencia» (de lo cuantitativo a lo cualitativo) en forma de nuevas estrategias, planteamientos y soluciones empresariales o de otros ámbitos, como el científico o el cuidado de la salud.

Correo electrónico: [rmorca@gmail.com](mailto:rmorca@gmail.com)

<https://doi.org/10.1016/j.angio.2018.05.001>

0003-3170/© 2018 SEACV. Publicado por Elsevier España, S.L.U. Todos los derechos reservados.

El concepto de *big data* no se refiere a una cantidad específica. Usualmente se utiliza cuando se habla en términos de *petabytes* y *exabytes* de datos. Analicemos primeramente en términos de *bytes*.

## Unidades de almacenamiento de información

*Byte* es la unidad de información de base utilizada en computación y telecomunicaciones, y que resulta equivalente a un conjunto ordenado de bits, generalmente 8 bits, por lo que en español también se le denomina octeto, del francés *octet*.

La unidad *byte* no tiene símbolo establecido internacionalmente, aunque en países anglosajones es frecuente representarlo como B, mientras que en los países francófonos con frecuencia se utiliza O (*octet*).

Bit es el acrónimo de *binary digit* (o sea de «dígito binario», en español denominado como bit y, en menor medida, como bitio). Un bit es un dígito del sistema de numeración binario. La capacidad de almacenamiento de una memoria digital también se mide en *bits*.

Lo usual es que un registro digital u otras memorias digitales vinculadas con la computación o con las telecomunicaciones tengan una capacidad de representación de información de, por ejemplo, 8, 16, 32, 64 bits, etc.; una memoria binaria tiene una capacidad efectiva de representación de un bit.

Mientras que en el sistema de numeración decimal se usan 10 dígitos (10 símbolos), en el binario se usan solo 2 dígitos, el 0 y el 1. Un bit o dígito binario puede representar uno de esos 2 valores: 0 o 1. Byte es el conjunto de 8 bits que recibe el tratamiento de una unidad y que constituye el mínimo elemento de memoria direccionable de un ordenador.

Las unidades utilizadas en múltiplos de bites se encuadran en 2 sistemas de medida el International Standardization Organization y el International Electrothecnic Comission (tabla 1).

## Conceptos clave asociados al *big data*

Los conceptos clave esenciales asociados al *big data* han dado en denominarse las 3V<sup>1</sup> (volumen, velocidad, variedad)

e inclusive las 6 V, añadiendo variabilidad, veracidad y valor a otros aspectos.

**Volumen.** El tamaño de la información.

**Velocidad.** Incluye tanto la media de velocidad en la que llegan los datos como también el tiempo en el que se debe actuar.

**Variabilidad.** Se refiere a la heterogeneidad de los datos, su representación y su semántica. Puede ser estructurada o no estructurada.

**Veracidad.** Tiene que ver con la precisión y la confianza de los datos que se manejan.

**Valor.** Se basa en el potencial para obtener información sobre las condiciones complejas que afectan a la salud de las personas y las poblaciones. Es importante destacar que *big data* tiene el potencial de ampliar enormemente el conocimiento para muchas condiciones clínicas en las que la recopilación de datos prospectivos y estructurados es prohibitiva en términos de tiempo y costo.

**Privacidad.** Los usuarios deben sentir confianza para suministrar la información. Las empresas deben tener procesos estrictos para su utilización. La protección de datos debe ser una prioridad.

**Complejidad.** Tiene que ver con transformar datos operativos en grandes plataformas de *big data* y la dificultad que implica gestionarlos en cualquier momento y desde cualquier lugar. La información puede ser estructurada (base de datos, transacciones, claves, columnas, registros) o no estructurada (correos electrónicos, informes, hojas de cálculo).

**¿De dónde procede toda esta información?**

## ¿De dónde procede toda esta información?

Los seres humanos estamos generando y almacenando constantemente información. La evolución tecnológica a la que se tiene acceso es una de las claves. El número de dispositivos móviles conectados a Internet a escala mundial es rápidamente creciente. La información que genera un *call center*, las notas de voz, correos electrónicos, documentos

Tabla 1 Múltiplos de bites

SI (decimal)		ISO/IEC 80000-13 (binario)	
Múltiplo (símbolo)	SI	Múltiplo (símbolo)	ISO/IEC
Kilobyte (kB)	10 <sup>3</sup>	Kibibyte (KiB)	2 <sup>10</sup>
Megabyte (MB)	10 <sup>6</sup>	Mebibyte (MiB)	2 <sup>20</sup>
Gigabyte (GB)	10 <sup>9</sup>	Gibibyte (GiB)	2 <sup>30</sup>
Terabyte (TB)	10 <sup>12</sup>	Tebibyte (TiB)	2 <sup>40</sup>
Petabyte (PB)	10 <sup>15</sup>	Pebibyte (PiB)	2 <sup>50</sup>
Exabyte (EB)	10 <sup>18</sup>	Exbibyte (EiB)	2 <sup>60</sup>
Zettabyte (ZB)	10 <sup>21</sup>	Zebibyte (ZiB)	2 <sup>70</sup>
Yottabyte (YB)	10 <sup>24</sup>	Yobibyte (YiB)	2 <sup>80</sup>

IEC: International Electrothecnic Comission; ISO: International Standarization Organization; SI: sistema internacional.

electrónicos o estudios médicos en todas sus dimensiones de texto e imagen, son algunos de los orígenes.

La industria, con datos de transacciones de clientes, proveedores u operaciones. Los gobiernos, con datos censales, impuestos, etc. Operaciones financieras realizadas en línea o por dispositivos móviles. Redes sociales, como Twitter, Facebook, LinkedIn, etc. Entre otras muchas.

Pero también las máquinas generan un crecimiento de la información (*machine-to-machine* [M2M]). Existen sensores digitales instalados en contenedores para determinar rutas en compañías de transporte. Sensores en medidores eléctricos para determinar consumos energéticos. Sensores interconectados en gran variedad de sectores industriales, de automoción, servicios, etc. Actualmente, la utilización del GPS es rutinaria y sus señales también contribuyen.

Los denominados *big transaction data*, que incluyen registros de facturación, o registros de llamadas en telecomunicación.

Por último, la biometría, en la que se incluye huella digital, escaneo de retina, reconocimiento facial o genética, entre otros. Muy importante a nivel de agencias de inteligencia y seguridad.

## Utilidades del *big data*

*El control de las comunicaciones.* El mundo digital y la interacción de los ciudadanos con nuevas formas de comunicación, a través de redes sociales, el almacenamiento en «la nube» (*cloud computing*) y nuevos métodos de difundir la información, son el sustrato sobre el que se ha generado el sistema denominado de espionaje y vigilancia.

Personalidades ya populares, como Julian Assange, editor y fundador de *Wikileaks*, Jacob Appelbaum, fundador de *Noisebridge*, o Edward Snowden, norteamericano tachado de héroe por algunos o traidor por otros, han difundido a la población general la situación actual al respecto.

Snowden, en 2013, ofreció información sobre el programa PRISM y multitud de documentos clasificados de la Agencia Nacional de Seguridad que se publicaron en *The Guardian* y *The Washington Post*, entre otros.

El impacto de sus revelaciones respecto al tratamiento y los usos de la información y la participación de las grandes compañías del mundo de las TIC, como Google, Microsoft, IBM, Facebook, Apple o Skype, aún se percibe en todo el mundo.

El interés de los gobiernos en esta materia puede encontrarse más oculto o ser tan explícito como en la Unión de Emiratos Árabes, en los que se acaba de nombrar a su primer ministro de Estado para Inteligencia Artificial, como parte de todo un proyecto gubernamental para su creación y desarrollo que incluye también la creación de una legislación específica.

*Aplicaciones en el mundo empresarial.* Conocer lo que se dice de una empresa y de sus competidores en las redes orienta para calibrar sus debilidades, fortalezas y mucho más. Conocer el comportamiento de los clientes y del mercado, sus preferencias y demandas, sirve para apoyar campañas de *marketing*, facilitar procedimientos de control de calidad, ayudar en una auditoría, mejorar el servicio al cliente, gestionar mejor el riesgo y el cumplimiento de la normativa.

Cualquier empresa que se precie requiere un analista de datos, ya sea de sus productos o de sus servicios. Con este fiel conocimiento del entorno es más capaz de organizar su estrategia y alcanzar las metas que se propone.

Las empresas más interesadas son las de telefonía, agricultura, eléctricas, la banca, grandes almacenes, las dedicadas al ocio y al cuidado de la salud.

## *Big data* en sanidad

Se genera gran cantidad de datos por los pacientes, los financiadores, los proveedores de servicios sanitarios y los profesionales. Su explotación por analogía con otros sectores nos configura una prometedora forma para practicar medicina personalizada, predictiva, preventiva y participativa (medicina 4P).

El sector sanitario también ha experimentado una revolución tecnológica, sobre todo derivada de cambiar sus antiguas estructuras y sistemas de atención al paciente para adaptarse a la realidad. Todos los profesionales sanitarios se enfrentan a cambios globales que no pueden dejar pasar y que están enfocados en la innovación, la tecnología, el diseño y la comunicación.

La investigación clínica está adaptando su metodología clásica de los ensayos clínicos<sup>2</sup> al *big data*. También lo están haciendo todas las «ómicas» (genómica, epigenómica, proteómica, metabolómica, microbiómica, etc.). Específicamente, algunas aplicaciones, como la gestión de la imagen, la génesis de biomarcadores, el diagnóstico precoz y el seguimiento de los pacientes, gracias a sensores que proporcionan datos biológicos o hábitos de comportamiento, están configurando otra realidad a la que también los cirujanos vasculares debemos adaptarnos.

## Beneficios del *big data* en el ámbito de la cirugía vascular

La existencia de este nuevo paradigma en sanidad puede tener también un amplio impacto entre nosotros. La obtención masiva de datos de muy diversa procedencia es la base para la utilización de la inteligencia artificial, de lo que se deriva la obtención de valor. Un mejor valor en atención médica requiere enfocarse hacia resultados.

Los diferentes niveles de analítica descriptiva (que pasa ahora), predictiva (que va a pasar) o prescriptiva (cuál es la mejor estrategia) plantean las bases para dar respuesta a las preguntas que nos hacemos como expertos en diferentes ámbitos.

En términos más concretos, el *big data* es útil para la toma de decisiones en cuanto a diagnóstico precoz, pronóstico y planificación de tratamientos.

Es también de gran utilidad en la gestión hospitalaria para diseño de protocolos, optimización de recursos y predicción de demanda. Existen experiencias con servicios sanitarios concretos, públicos y privados. Estamos evolucionando hacia una salud basada en datos.

La inteligencia artificial se está comenzando a aplicar al análisis de texto de las historias clínicas y estos datos están llevando a generar nuevo conocimiento médico basado en nuevas evidencias.

El análisis de la práctica clínica real nos permitirá optimizar los estándares de calidad, en cada proceso clínico.

Mediante algoritmos *machine learning* se puede predecir con precisión, por ejemplo, la tasa de asistencia a urgencias, según la influencia de determinados eventos sociales coincidentes.

La investigación puede impulsarse acelerando el reclutamiento de pacientes de manera exponencial. Sirva de ejemplo la petición de Obama para la investigación del genoma humano en 2015.

Algunas sociedades científicas también están adaptándose a esta nueva realidad como la radiología<sup>3</sup> o la Sociedad Española de Cardiología, que investiga la relación entre salud bucodental y enfermedades cardiovasculares.

La implementación de ciertos sistemas puede ser capaz de realizar tareas, como cita *on-line*, seguimiento de pacientes complejos, seguimiento de la adherencia terapéutica, fomento de los hábitos de vida saludables y soportes a nivel de *call center*. Inclusive de «asistentes virtuales». Ya existen experiencias en este sentido en Reino Unido para para la artritis.

Teniendo en cuenta todos estos profundos cambios, debiéramos además de ilustrarnos en las peculiaridades de todo este ámbito, conseguir tener presencia, comunicación interprofesional y actualización (investigación e innovación) inmediata.

Proyectos como «*cardioTuit*»<sup>4</sup>, impulsado por la Sociedad Española de Cardiología, son un ejemplo de cómo suscitar el conocimiento y la inmersión en redes sociales como profesionales. Quizás adquiriendo visibilidad como cirujanos vasculares tengamos más peso en la comunidad científica, social y ante la Administración.

Con relación a nuestra especialidad, ya se encuentran disponibles aplicaciones para el autocontrol de enfermedades, algunas de elevada prevalencia, como la diabetes mellitus. A través de las aplicaciones generadas, algunas por la propia industria, se consigue orientación en la confección

de dietas, consumo calórico y tratamiento específico de control de la glucemia.

Otra de las grandes líneas de actuación en angiología y cirugía vascular podría ser la creación de un *AngioTuit* y, cómo no, la transformación de los registros de actividad.

Lo cual podría constituir una revolución de los clásicos conceptos por los cuales los efectuamos y explotamos. Este nuevo paradigma podían venir a transformarlos, utilizando diferentes entradas, entre las que podrían contemplarse los datos de facturación, reflejando las peculiaridades de los diferentes dispositivos empleados, el registro oficial de los pacientes en los centros públicos o privados, los datos censales y epidemiológicos o los cambios en los estilos de vida y su impacto en la calidad percibida por el propio paciente y analizados desde las entradas generadas por aplicaciones de autoconsumo, que sin duda aumentan el empoderamiento del paciente y ayudan a los profesionales a realizar una medicina preventiva y participativa.

## Conflicto de intereses

Los autores declaran no tener ningún conflicto de intereses.

## Bibliografía

1. McCue ME, McCoy AM. The scope of big data in one medicine: Unprecedented opportunities and challenges. *Front Vet Sci.* 2017;4:194, <http://dx.doi.org/10.3389/fvets.2017.00194>
2. Mayo CS, Matuszak MM, Ten RK, Haken. Big data in designing clinical trials: Opportunities and challenges. *Front Oncol.* 2017;7:187, <http://dx.doi.org/10.3389/fonc.2017.00187>
3. Kharat AT, Singhal S. A peek into the future of radiology using big data applications. *Indian J Radiol Imaging.* 2017;27:241–8.
4. Fernandez F, Gómez de Diego JJ. # *CardioTuit*. Aplicaciones profesionales de Twitter en cardiología. *Sociedad Española de Cardiología*; 2014.