

TRATAMIENTO ESTADÍSTICO DE BIG DATA: UN CAMBIO DE PARADIGMA TECNOLÓGICO EN LA UTILIZACIÓN DE LA INFORMACIÓN

ALFONSO JOSÉ LÓPEZ RIVERO

Universidad Pontificia de Salamanca. Salamanca. España
ajlopezri@upsa.es

Resumen: Los macrodatos constituyen el mayor recurso natural de la nueva era, similar al papel que han jugado la máquina de vapor, la electricidad o el petróleo en los siglos precedentes. Todos somos generadores de información, que nos puede ayudar o nos pueden despojar de todo tipo de privacidad. Además, los gobiernos bajo las directrices de transparencia y gobierno abierto ponen a disposición de los ciudadanos los datos que gestionan, lo que se conoce como *open data*. Aquellos que se pueden reutilizar se engloban bajo la denominación de *big data*. Estos datos sin procesar no generan conocimiento, es necesario validarlos y analizarlos. Este nuevo paradigma los transforma en información y ésta en valor.

Se estima que, en los próximos cinco años, se generarán \$13 billones de beneficios en los países pertenecientes al G20, debido a nuevos modelos de negocio basados en el análisis tecnológico dirigido a la reutilización de los datos para convertirlos en información. Estas cifras evidencian un nuevo motor de desarrollo económico de valor inestimable, que repercute directa o indirectamente en la calidad de vida de los ciudadanos.

Palabras clave: Valor de datos, análisis de macrodatos, economía basada en datos.

STATISTICAL ANALYSIS OF BIG DATA: A TECHNOLOGICAL PARADIGM SHIFT IN THE USE OF INFORMATION

Abstract: Macrodata are the greatest natural resource of the new era, similar to the role played by the steam engine, electricity or oil in the preceding centuries. We all generate information that can help us or be stripped of all privacy. In addition, governments under the guidelines of transparency and open government made available to the public the data they manage, what is known as open data. Those that can be reused are grouped under the name of big data. These raw data do not generate knowledge, validate and analyze necessary. This new paradigm in information and transforms it into value.

It is estimated that in the next five years, will generate \$ 13 billion in profits in countries belonging to the G20, because of new business models based on technological analysis aimed at the reuse of data into information. These figures show a new engine of economic development of inestimable value, which directly or indirectly affects the quality of life of citizens.

Key Words: value data, macro data analysis, data-driven economy.

El profesor Arranz en su artículo, “El hombre de cristal. La ambigua gestión del conocimiento humano”, parte de la reflexión “el conocimiento ha sido la más valiosa de nuestras materias primas”, y diserta sobre las prótesis que permiten gestionar el conocimiento adquiriéndose mediante el tratamiento de datos, ahondando en el problema que supone identificar y disponer de muchos datos frente al conocimiento, “*Google* no dispensa del esfuerzo de pensar” (Arranz Rodrigo, 2013).

Este trabajo, partiendo de la idea del profesor Arranz, analiza la creación de valor aplicando la tecnología y el tratamiento estadístico a las fuentes de macrodatos disponibles. Esta información se obtiene de las interacciones que realizamos las personas cuando interactuamos en internet, cuando accedemos a una red social o de los que facilitamos a las administraciones públicas. Además, hay que tener en cuenta los datos que provienen de los sensores que utilizamos frecuentemente, directamente como los GPS o indirectamente como el contador de la luz.

La organización de todos los conjuntos de datos, macrodatos, se conoce como *big data* y están disponibles en una determinada estructura y con unas características conocidas como las “Vs”. Si estos datos carecen de propiedad intelectual, se conocen como datos abiertos. Las nuevas políticas de transparencia de los gobiernos facilitan a los ciudadanos la información de la que disponen en este formato.

Estas circunstancias, disponer de una materia prima, datos libres de derechos de autor y de tecnología para analizarla y convertirla en valor, están abriendo oportunidades en la generación de una nueva economía de mercado en los países industrializados que los gobiernos ya empiezan a incluir como un nuevo elemento de crecimiento económico. Supone el nacimiento de distintas estrategias que pueden dar nuevas soluciones a la sociedad y llegar a nuevos mercados deslocalizados y globales.

La idea de utilizar los datos internos de una empresa para analizarlos y tomar decisiones que permitan ser más competitivos se aplicaba ya en el *business analytics* hace más de 60 años. La diferencia con el nuevo paradigma radica en que antes, fundamentalmente, se usaban datos externos a las compañías o a los usuarios en general sin conocer en la mayoría de los casos ni su procedencia ni su validez.

1. TRATAMIENTO MASIVO DE DATOS: *BIG DATA*

El concepto de Big Data, o tratamiento masivo de datos, surge del uso de la tecnología y también del cambio de hábitos en las interacciones sociales, puesto que los macrodatos aparecen en cada uso tecnológico que se realiza. Se conoce como modelo de las V's: volumen, variedad, velocidad, veracidad, valor de los datos y visualización, puntos que responden a cada uno de los interrogantes que nos surgen al estudiar y analizar los macrodatos.

Volumen de datos: las magnitudes están cambiando ya no se expresa en gigabytes si no de petabytes o de exabytes¹. El aumento masivo de datos se genera de diversas formas: las empresas guardan grandes cantidades de datos, reuniendo información de sus clientes, proveedores, o de las operaciones que realizan, de la misma manera que hacen todas las administraciones públicas con todos los datos que gestionan. En los países industrializados, se generan bases de datos, enormes, que guardan información sobre la población y todos los registros que realizan. A todos estos datos hay que añadirles transacciones financieras o los movimientos de información que se realizan por dispositivos móviles, o en las redes sociales. Considerando únicamente todas y cada una de las actividades que la mayoría de nosotros realizamos varias veces al día con nuestros dispositivos móviles generan alrededor de 2.5 quintillones² de bytes diariamente en el mundo.

De acuerdo con un estudio realizado por Cisco³, publicado en 2012, la cantidad de tráfico de datos móviles crecerá a una tasa anual del 78%. Además, afirma que el número de dispositivos móviles conectados a internet excederá al número de habitantes en el planeta. Por otra parte, las Naciones Unidas proyectan que la población mundial alcanzará los 7.5 billones para el 2016⁴, de tal modo que habrá cerca de 18.9 billones de dispositivos conectados a la red a escala mundial. Esto conllevaría que el tráfico global de datos móviles alcanzará 130 exabytes anuales. Este volumen de tráfico previsto para 2016, de 33 billones de DVDs anuales, puestos uno a continuación de otro equivaldría a la distancia de ida y vuelta a la luna.

No solo los seres humanos contribuyen a este crecimiento enorme de información, existe también la comunicación llamada máquina a máquina, M2M, cuyo valor en la creación de grandes cantidades de datos también es considerable.

1 1 Exabyte = 10^{18} = 1.000.000.000.000.000.000 de bytes.

1 Petabyte = 10^{15} = 1.000.000.000.000.000 de bytes

2 1 quintillón = 10^{30} = 1,000,000,000,000,000,000,000,000,000,000

3 <http://www.cisco.com/web/LA/soluciones/strategy/education/connection/articles/edition4/article4.html>

4 <http://www.un.org/en/development/desa/population/>

Como ejemplo, cabe señalar los nuevos contadores de la luz que disponen de sensores para medir el consumo de energía a intervalos regulares de tiempo y cuya información es enviada a las máquinas de las compañías proveedoras de energía. En el sector de la salud, se están transformando además los dispositivos médicos y de salud personales. Estos hábitos provocarán fuertes cambios en los sistemas mundiales de salud y se estima que en Estados Unidos existirán 386 millones de conexiones en 2023, seguidos por Europa y Asia Pacífico. Con este crecimiento significativo de dispositivos y sensores conectados, la industria de la salud M2M generará 90.9 millones de dólares en ingresos totales en 2023 con un ritmo de crecimiento del 17,2 %.

Variedad: los datos se pueden obtener de diversas maneras en todo el mundo, por dispositivos móviles, audio, video, sistemas GPS, sensores digitales, automóviles, medidores eléctricos, veletas, anemómetros, etc., que pueden medir y comunicar el posicionamiento, movimiento, vibración, temperatura, humedad y hasta los cambios químicos que sufre el aire. La Comisión de Estadística de Naciones Unidas en el informe, *Macrodatos y modernización de los sistemas estadísticos* de marzo de 2014, clasifica la variedad de los macrodatos en seis tipos⁵:

- Fuentes que surgen de la administración de un programa, gubernamental o no, como los historiales médicos electrónicos, las visitas hospitalarias, los archivos de seguros, los archivos bancarios y los bancos de alimentos.
- Fuentes comerciales o transaccionales derivadas de las transacciones entre dos entidades, como las transacciones con tarjetas de crédito y las transacciones en línea (incluso las realizadas desde dispositivos móviles).
- Fuentes procedentes de las redes de sensores, como las imágenes obtenidas por satélite, los sensores de las carreteras y los sensores climáticos.
- Fuentes provenientes de los dispositivos de seguimiento, como los datos de seguimiento obtenidos de los teléfonos móviles y el sistema mundial de determinación de la posición (GPS).
- Fuentes de datos derivados de comportamiento, como las búsquedas en línea (de un producto, un servicio o cualquier otro tipo de información) y el visionado de páginas en línea.
- Fuentes de datos de opinión, como los comentarios en las redes sociales.

Velocidad de respuesta: se refiere a la rapidez de generación de los datos, y supone un reto no solo como capacidad para almacenarlos, sino también para procesar datos y dar respuestas en tiempo real. Cuando se realiza una consulta en Google, en una red social o cuando compramos un billete de avión por internet,

5 <http://unstats.un.org/unsd/statcom/doc14/2014-11-BigData-S.pdf>

estamos realizando consultas a fuentes de datos que requieren de una velocidad de respuesta suficientemente rápida a la petición requerida.

Veracidad o confianza: cualquier actividad pueda ser traducida en dato. Se tienen que obtener datos de calidad eliminando las características inherentes a su obtención o el sitio del que provienen o donde se almacenan, para obtener una correcta toma de decisiones.

Valor de los datos: la importancia de la información para el negocio, conocer qué datos son los que se deben analizar es fundamental. De manera, que ya se empieza a hablar del científico de datos, un profesional con perfil científico, tecnológico y con visión de negocio. La cadena de valor de *big data* proporciona un marco que examina cómo reunir datos dispersos de forma organizada y crear información de valor que pueda servir para la toma de decisiones a nivel empresarial, incluyendo así cuatro grandes fases: la generación y adquisición de los datos, su procesado y análisis, la posterior curación y almacenamiento de los mismos y, finalmente, la creación de servicios asociados (Antolín Fernández, *et al.*, 2015).

Visualización: con la gran cantidad de macrodatos es necesario disponer de formas más accesibles y fáciles de leerlos y analizar resultados complejos de forma sencilla. La visualización interactiva permite a los usuarios personalizar e interactuar con los resultados, estas representaciones visuales algunas veces tendrán que ir acompañadas de una breve narrativa para proporcionar contexto y sentido. En palabras de (Antolín Fernández, *et al.*, 2015):

La visualización de los datos es vital si se quiere un consumo efectivo de Big Data. Los informes que se generan de la analítica de datos pueden ser documentos complejos. Para que los usuarios interpreten rápida y correctamente toda esta cantidad de documentos, se deben diseñar cuidadosamente la visualización digital y la presentación. Las interfaces de información deben humanizarse, es decir, responder a las necesidades y capacidad humanas.

El informe anual del BBVA⁶, afirma sobre las tendencias de *big data*:

Big Data ya no es una promesa ni una tendencia, está aquí y está provocando cambios profundos en diversas industrias. Desde el punto de vista tecnológico ya existen sectores empresariales que han adoptado de forma masiva proyectos y productos. El análisis de todos los datos disponibles está convirtiéndose en un elemento de disrupción. Así como internet es un factor de desintermediación que está afectando a muchas cadenas de valor, el análisis de información en grandes volúmenes, de diversas fuentes, a gran velocidad y con una flexibilidad sin precedentes puede suponer un factor diferencial para aquellos que decidan adoptarlo.

6 <https://www.centrodeinnovacionbbva.com/innovation-edge/big-data>

2. VALOR DEL TRATAMIENTO TECNOLÓGICO DE LOS DATOS

Un informe de la entidad independiente McKinsey Global Institute, (The Mckinsey Global Institute, 2013) estima que, en los próximos cinco años, se generarán \$13 billones de beneficios en los países pertenecientes al G20 (<https://g20.org/>, s.f.), debidos al modelo de negocio basado en la generación de la actividad económica de las empresas que tienen como cartera de productos la reutilización de los datos abiertos. Estas cifras evidencian que la utilización de la información del sector público se ha convertido en un motor para el desarrollo económico y tiene un valor inestimable que repercute directa o indirectamente en la calidad de vida de los ciudadanos. De una forma directa se tienen aplicaciones de soluciones big data, basadas en móviles o en la red, y de forma indirecta además de un nuevo yacimiento de empleo, supone la captación de nuevos mercados internacionales.

Las políticas de datos abiertos ofrecen un nuevo camino para la reforma económica y el crecimiento de la productividad. Aprovechando las posibilidades de las nuevas tecnologías, permiten hacer un mejor uso de los recursos y promover el objetivo de crecimiento del G20 del 2%. Los informes realizados sugieren que la aplicación de políticas de datos abiertos podrían aumentar el PIB acumulado de los países en torno al 1,1%, siendo la previsión de crecimiento hasta el 2018 del 2% (Gruen, *et al.*, 2014).

A medida que la economía y la sociedad se basan cada vez más en el conocimiento, los datos son activos fundamentales y su reutilización son parte fundamental del crecimiento y desarrollo económico. La puesta en valor de la Información del Sector Público, ISP, y la reutilización de los datos en combinación con las nuevas tecnologías suponen una mejora en la creación de valor económico y social. Según (Gruen, *et al.*, 2014), favorecen fundamentalmente la reducción de los costos en la prestación de servicios, tanto para el gobierno como para el sector privado, la creación y perfeccionamiento de la calidad de nuevos servicios y suponen avances en la transparencia de la gestión de gobierno, mediante la mejora de la rendición de cuentas y la participación del ciudadano generando una mayor confianza en los gobiernos.

Para obtener utilidad de los datos masivos derivados del gobierno abierto, es imprescindible transformarlos en información aprovechable a través de las técnicas utilizadas en el *business intelligence*, basadas en conjuntos de datos limitados y que se preguntan ¿qué pasó?, dando lugar al análisis predictivo, y a la generación de documentos. El siguiente paso es trabajar con datos ilimitados, *Big Data*, que analiza cantidades de datos aparentemente ilimitados y que puede determinar lo que es probable que suceda. Es aquí donde está el verdadero poder del *Big Data*, porque esa información, analizada y cruzada, genera conocimiento

que puede resultar imprescindible para mejorar la toma de decisiones. La gestión inteligente de la información; tanto la estructurada –la que proviene de transacciones y de datos con un formato concreto–, como la no estructurada –social media, navegación o la basada en la localización–, ofrece un gran potencial para mejorar muchos procesos de toma de decisiones estratégicas y avanzar en innovación, productividad y competitividad. Un informe de Oracle, (ORACLE, 20 de febrero de 2013) advierte que, probablemente, en la comercialización masiva de productos, si se aplicasen técnicas Big Data, se podrían incrementar los márgenes en un 60% y los costes de fabricación y mantenimiento de las empresas se verían reducidos a casi la mitad.

La inversión en tecnologías de grandes datos sigue creciendo, según un estudio de diciembre de 2014 realizado por Gartner, Inc.⁷. En este se destaca que el 73% de los encuestados han invertido o planean invertir en grandes volúmenes de datos en los próximos dos años, frente al 64% que lo hacía en 2013. La encuesta también indica que las organizaciones están empezando a bajar la valla de su gran inversión de datos –el número de organizaciones que indican que no tenían planes para la inversión de grandes datos se redujo de 31% en 2013 al 24% en 2014 (Stamford, 14).

El aumento de datos genera oportunidades de negocio. En el universo digital se espera un crecimiento del 40% hasta el 2020 (Turner, et al., 2014). Esta expansión de dispositivos inteligentes conectado a Internet, implica un aumento de personas, científicos de datos, y de empresas que trabajen en el análisis predictivo de los datos. Al igual que el universo físico, el universo digital duplica su tamaño cada dos años, y en 2020 el universo digital llegará a 44 zettabytes, o su equivalente en las unidades actuales 44 billones de gigabytes.

3. POLÍTICA DE GOBIERNO ABIERTO

Desde que Obama promulgara el “Memorándum de Transparencia y Gobierno Abierto”, aprobada por el gobierno en el año 2009, se ha abierto un camino en la transparencia de los gobiernos que ha llevado a lo que hoy se conoce como *open data* (Observatorio Nacional de las Telecomunicaciones y de la Sociedad de la Información (ONTSI, 2013). El tratamiento de esta información y, en particular, la aplicación de la metodología estadística a la gran cantidad de datos, *big data*, ha abierto una vía a la puesta en valor de la economía basada en la red, generando aplicaciones, productos o servicios de valor añadido destinados a terceros.

7 <http://www.gartner.com/technology/topics/big-data.jsp>

La aparición del *open data* en Europa surge de las políticas europeas que impulsan el gobierno abierto a través de la “Directiva 2003/98/CE del Parlamento Europeo y del Consejo” en la que se establece la reutilización de la información del sector público. Su finalidad era facilitar la reutilización de la Información del Sector Público, ISP, en toda la Unión Europea, pactando las condiciones básicas de reutilización y eliminando los principales obstáculos para la reutilización en el mercado interior. A partir de esta directriz europea, los países miembros han fomentado la transparencia de sus administraciones públicas a través de la aprobación de leyes que permiten el acceso a la información del sector público.

Los países pertenecientes a la Organización para la Cooperación y Desarrollo Económico, (OECD), trabajan conjuntamente para compartir experiencias y buscar soluciones a los problemas comunes. Pretenden entender qué es lo que conduce al cambio económico, social y ambiental, midiendo la productividad y los flujos globales del comercio y la inversión, analizando y comparando datos para realizar pronósticos de tendencias y fijando estándares internacionales dentro de un amplio rango de temas de políticas públicas. La OECD publica en el año 2007 los principios y directrices para el acceso a los datos de la investigación financiada con sus fondos públicos; la creación de un gobierno abierto que colabore con la sociedad en la generación de valor público, la producción de información de servicios públicos para lograr una administración eficaz, eficiente, innovadora, flexible, transparente, profesional y responsable, entre otras características, (Organisation for Economic Co-operation and Development, -OECD-, 2007).

La puesta en valor de la Información del Sector Público, ha abierto una nueva vía a la transparencia y a la comunicación con los ciudadanos. Si bien el reto continúa siendo el acceso a todos los datos disponibles. Las Tecnologías de la Información y la Comunicación han facilitado las posibilidades de acceso a una cantidad ingente de datos, *Big Data*, abriendo nuevas vías en la reutilización de toda la información pública. El acceso a la información en sí mismo no supondría un valor añadido si no se tratan los datos que se facilitan por las administraciones públicas. El empleo de herramientas estadísticas al análisis y procesamiento de toda esta información ha permitido el nacimiento de empresas que están desarrollando modelos de negocio basados en la creación de aplicaciones, productos y servicios de valor añadido a partir de la información del sector público.

GENERANDO VALOR DE EMPRESA CON EL TRATAMIENTO DE LA INFORMACIÓN DE GRANDES DATOS

En España, según el “Estudio de Caracterización del Sector Economía Digital” del Observatorio Nacional de las Telecomunicaciones y la Sociedad de la

Información, ONTSI (Observatorio Nacional de las Telecomunicaciones y de la Sociedad de la Información (ONTSI, 2013), los casos de éxito en el uso de la reutilización de la información pública generan un volumen de negocio anual de entre 550 y 650 millones de euros y emplean directamente a cerca de 5.500 trabajadores. En esta línea, España también ocupa una posición de liderazgo en cuanto a la calidad de los servicios de la Administración electrónica. En ese sentido, el último análisis realizado por la Comisión Europea señala que España alcanzó un 98% de madurez en sus servicios de Administración electrónica, figurando entre los diez países más avanzados dentro de la UE.

De acuerdo al informe de la ONTSI, el negocio de los contenidos digitales se encuentra en España al mismo nivel que en otros países. De los 8.000 millones de euros que produjo la industria digital en 2009, la publicidad a través de internet generó unos 650 millones de euros, una cifra de negocio similar al sector que basa su negocio en la utilización del *open data*. La actividad de las empresas con base en el tratamiento de datos abiertos, consideradas en el estudio, representa entre un 35 y 40% de la actividad total de las empresas digitales, alcanzando un volumen de negocio global superior a los 1.600 millones de euros.

La internacionalización de los clientes de las empresas españolas derivadas de esta actividad económica, según el informe de la ONTSI, se ha incrementado, captando clientes de la Unión Europea y de fuera de ella. De los datos del estudio, se obtiene que los de fuera de España representan un 45,7%, pertenecientes a empresas de la Unión Europea, y el 20% vienen de fuera de la UE. En general, las empresas basadas en *Big Data* tienen un alto nivel tecnológico; y la innovación en el campo del análisis, tratamiento y presentación de la información reutilizada es una de sus principales características.

Según la ONTSI, la inversión media anual destinada a tecnologías en España es de 915.000 euros, y una parte importante de esta se destina a la mejora e innovación en la gestión, tratamiento y el análisis de los datos. Con estas cifras se puede considerar que el sector presenta grandes oportunidades gracias, entre otros factores, al creciente volumen de información del sector público puesto en valor y al potencial en el desarrollo de nuevos servicios que permite la reutilización de estos datos, nacionales o extranjeros. La combinación de estos factores posibilita además la captación de nuevos mercados potenciando la internacionalización de los clientes.

Con el fin de divulgar y aumentar la visibilidad de las iniciativas de reutilización de los datos facilitados por las administraciones públicas en Estados Unidos, el portal *datos.gov* ofrece una sección donde se recogen una veintena de casos de éxitos de compañías reutilizando el *open data* a escala estatal. Este escaparate muestra un listado empresarial, clasificado por sectores de actividad –finanzas,

servicio, salud, energía y educación–, donde se detalla la línea de negocio, la ubicación de la compañía, sus datos financieros y el número de empleados.

4. ANÁLISIS DE MACRODATOS

Los datos analizados generan conocimiento y este aporta valor. Desde los años 60 este proceso se ha conocido como inteligencia de negocios, convirtiéndose en un paradigma que transformaba los datos en modelo de negocio para poder tomar decisiones, cómo, ¿cuál es el producto que mejor se puede vender?

Si nos preguntamos sobre los condicionantes que hacen que un producto sea el que mejor se vende, sus características intrínsecas, o los condicionantes del entorno, o las preferencias de los potenciales consumidores, de manera tradicional la información se adquiría por el “boca a boca”, actualmente esa difusión se realiza a través de internet o se comparten comentarios en las redes sociales. Todos estos sensores generan innumerables datos, macrodatos, con informaciones complementarias que nos puede permitir hacer entender el ¿por qué? de las cosas, y además, en muchas situaciones, poder predecir y anticiparnos a los acontecimientos.

El análisis de macrodatos requiere las cuatro fases que se utilizan en el análisis estadístico y econométrico: en una primera fase, se detalla lo que está ocurriendo en ese momento y corresponde a una *descripción* de los acontecimientos. En la segunda fase, se quieren obtener respuestas a lo que originó la causa, se está en la etapa de *diagnóstico*. Si lo que se pretende es anticiparse a los acontecimientos estaríamos en el momento de la *predicción*, mientras que si lo que se busca es analizar los datos para poder anticiparse a los acontecimientos estaríamos en la última etapa, la *prescriptiva*.

Los modelos descriptivos y de diagnóstico prescriptivo analizan las relaciones que se establecen en los datos para poder clasificarlos en distintos grupos, identificando las diferentes relaciones existentes entre las características de los datos.

La fase de análisis predictivo utiliza fundamentalmente algoritmos de minería de datos y se basa en el análisis del conjunto de los datos actuales e históricos para poder realizar predicciones sobre futuros eventos. Estas afirmaciones no se pueden considerar como afirmaciones absolutas, puesto que tienen una validez en intervalos próximos de tiempo con una probabilidad de que dichos eventos puedan ocurrir. Normalmente, estos modelos se utilizan para determinar riesgos potenciales o vías hacia las cuales se pueden encaminar las tomas de decisiones, así por ejemplo en un análisis predictivo en el campo de la medicina se obtienen orientaciones en la toma de decisiones de cuáles pueden ser los fármacos que

mejor se adecúan en la evolución de una determinada patología. En el mundo de los negocios, se puede determinar el riesgo que se puede asumir ante una determinada situación económica.

La toma de decisiones basada en el análisis de los datos se fundamenta en la clasificación que se puede realizar de ellos como materia prima. Estos se pueden agrupar considerando los lugares de los que provienen, lo que se conoce como *fuentes de datos*; o bien de acuerdo a la estructura que tienen los archivos donde se almacenan en formato digital, denominándose esta clasificación *tipología de los datos*, reflejada en la Tabla 1

Fuentes de Datos	Externos	<ul style="list-style-type: none"> • Teléfonos móviles. • Historiales de viajes • 	<ul style="list-style-type: none"> • Redes sociales • Sensores externos • ...
	Internos	<ul style="list-style-type: none"> • Recurso Humanos • Ventas y datos financieros • 	<ul style="list-style-type: none"> • Foros de participación. • Documentos de texto • ...
		<i>Estructurados</i>	<i>Desestructurados</i>
Tipos de Datos			

Tabla 1: *Clasificación de los datos, por origen y por almacenamiento.*

El análisis de los datos que provienen de fuentes internas y están estructurados son los que conocemos de antemano. Su análisis permite realizar una prospección hacia los datos externos estructurados obteniéndose por este método información del exterior. El análisis de datos internos que se encuentran almacenados de una forma desestructurada permite realizar análisis de aprendizaje para poder trabajar en el área de datos externos desestructurados, siendo esta área la mejor oportunidad para recoger información exterior al sistema.

Joyanes en la publicación, *Análisis de grandes volúmenes de datos en organizaciones* (Joyanes Aguilar, 2013), viene a confirmar que el trabajo conjunto del *Big Data* con el análisis de negocio ha sido perfecta, produciendo modelos que se ajustan mejor al planteamiento que se pretende resolver, y abriendo camino a nuevos análisis que se enfrentan con problemas complejos como el ajuste de modelos a toma de decisiones en tiempo real. Esta unión se ha ido transformando en el almacenamiento de todos los datos internos y externos, estructurados y desestructurados, *Big Data*, para el análisis sistemático y automatizado. Este nuevo paradigma es en el que trabajan compañías como Google y que permite hacer afirmaciones tan arriesgadas como: “serán ellos los que ofrezcan determinados

servicios, ya que conocerán literalmente la oferta y la demanda antes de que se produzca”.

Esta afirmación tan arriesgada no contempla las correlaciones causales y se puede incurrir en el riesgo de obtener asociaciones estadísticamente robustas pero que se obtienen de la causalidad. El análisis de macrodatos puede aportar nuevas hipótesis y posibles correlaciones entre variables e indicar caminos hacia dónde iniciar un análisis, pero para crear modelos siempre será necesario un conocimiento mayor del problema, que solo poseen los investigadores en la materia.

Lo que conocemos como *Big Data* y su análisis ha supuesto un nuevo paradigma analítico que aporta valor a los datos. Difiere de planteamientos precedentes en que no se conocen las preguntas que se realizan a priori y que se puede disponer de ellos, en muchos casos, libres de propiedad intelectual, *open data*, o bien por las políticas de *gobierno abierto*. En ambos casos, es una materia prima de la que se puede disponer libremente. Este factor está suponiendo, además de aportar soluciones a problemas que antes no los tenían, una nueva forma de que el valor de los datos genere nuevos modelos de negocio; como se está poniendo de manifiesto en la nueva economía de mercado de los países industrializados, tal y como se reflejan en sus previsiones de crecimiento para los próximos años.

REFERENCIAS BIBLIOGRÁFICAS

- ANTOLÍN FERNÁNDEZ, M., RICAU GONZÁLEZ, F. y SÁEZ DOMINGO, D., (2015), *Big Data: El valor de los Datos. Estado actual y tendencias del Big Data como nuevo activo en la economía europea*. Valencia: Colegio Oficial de Ingenieros en Informática de la Comunidad Valenciana (COIICV).
- ARRANZ RODRIGO, M., (2013), “El hombre de cristal. La ambigua gestión del conocimiento humano”. En: *Estudio Agustiniano*, XLVIII/1, 141-152.
- GRUEN, N., HOUGHTON, J. and TOOTH, R., (2014). <http://apo.org.au>. [En línea], Available at: <http://apo.org.au/research/open-business-how-open-data-can-help-achieve-g20-growth-target> [Último acceso: 20 abril 2015].
- JOYANES AGUILAR, L., (2013), *Big Data. Análisis de grandes volúmenes de datos en organizaciones*. Barcelona: Marcombo, S.A.
- OBSERVATORIO NACIONAL DE LAS TELECOMUNICACIONES Y DE LA SOCIEDAD DE LA INFORMACIÓN (ONTSI), (2013), *Estudio sobre Objetivos, Estrategias y Actuaciones Nacionales e Internacionales en Materia de Gobierno Abierto*. Madrid: Ministerio de Industria, Energía y Turismo.
- ORACLE, 20 de febrero de 2013, *Big Data y su impacto en el negocio*, s.l.: Oracle.

- ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, -OECD, (2007), *OECD Principles and Guidelines for Access to Research Data from Public Funding*, Paris, France: OECD.
- PIATTINI VELTHUIS, M. G., GARCÍA RUBIO, F. O., GARCÍA RODRÍGUEZ DE GUZMÁN, I. y PINO, F., (2015). *Calidad de Sistemas de Información*. (3ª ed.). Madrid: RA-MA.
- STAMFORD, Conn., 14. <http://www.gartner.com>. [En línea] Available at: <http://www.gartner.com/newsroom/id/2848718>
- THE MCKINSEY GLOBAL INSTITUTE, (2013), <http://www.mckinsey.com>. [En línea] Available at: http://www.mckinsey.com/insights/business_technology/open_data_unlocking_innovation_and_performance_with_liquid_information [Último acceso: 10 marzo 2015].
- TURNER, V., GANTZ, J. F., REINSEL, D. and MINTON, S., (2014). *The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things*. s.l., s.n.
- <http://www.cisco.com/web/LA/soluciones/strategy/education/connection/articles/edition4/article4.html>
- <http://www.un.org/en/development/desa/population/>
- <http://unstats.un.org/unsd/statcom/doc14/2014-11-BigData-S.pdf>
- <https://www.centrodeinnovacionbbva.com/innovation-edge/big-data>
- <http://www.gartner.com/technology/topics/big-data.jsp>
- <https://g20.org/>, s.f. <https://g20.org/>. [En línea] Available at: <https://g20.org/>