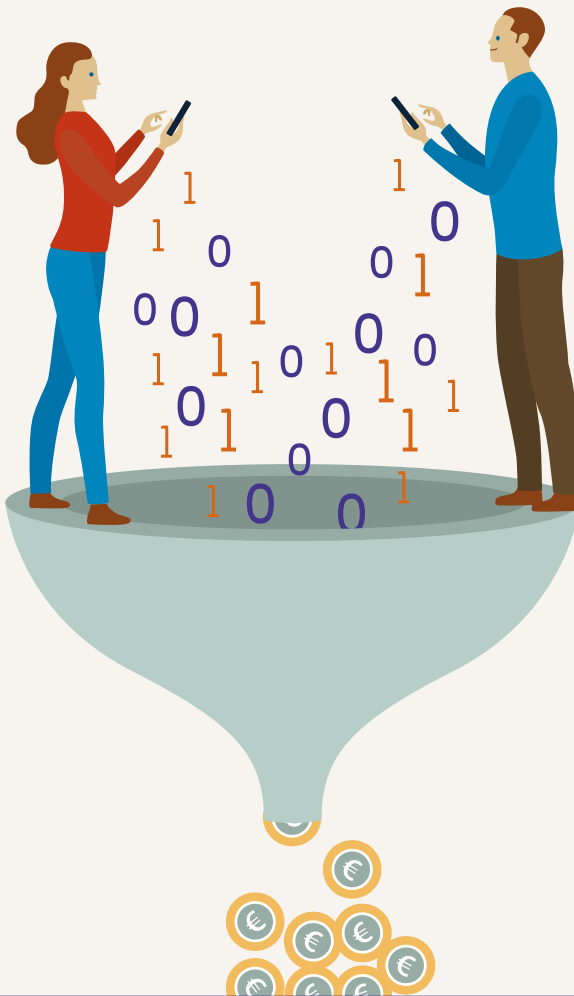


Economía de los Datos

Riqueza 4.0

Emilio Ontiveros (dir.)

Verónica López Sabater (coord.)



Economía de los Datos

Director:
Emilio Ontiveros

Coordinadora:
Verónica López Sabater



Ariel



Telefónica
FUNDACIÓN

Esta obra ha sido editada por Ariel y Fundación Telefónica, en colaboración con Editorial Planeta, que no comparten necesariamente los contenidos expresados en ella. Dichos contenidos son responsabilidad exclusiva de sus autores.

© **Fundación Telefónica, 2017**

Gran Vía, 28
28013 Madrid (España)

© **Editorial Ariel, S. A., 2017**

Avda. Diagonal, 662-664
08034 Barcelona (España)

© de los textos: Fundación Telefónica
© de la ilustración de cubierta: Jmugica
© del diseño de cubierta: LACASTA

Coordinación editorial de Fundación Telefónica: Rosa María Sáinz Peña
Primera edición: diciembre de 2017

El presente monográfico se publica bajo una licencia Creative Commons del tipo: Reconocimiento - Compartir Igual



Esta obra se puede descargar de forma libre y gratuita en:
www.fundaciontelefonica.com/publicaciones

ISBN: 978-84-08-18553-6

Depósito legal: B. 1.952-2018
Impresión y encuadernación: UNIGRAF
Impreso en España – Printed in Spain

El papel utilizado para la impresión de este libro es cien por cien libre de cloro y está calificado como **papel ecológico**.

Índice

Prólogo — 7
Sobre los autores — 9
Resumen ejecutivo — 11

Parte I - Introducción

1

Introducción: la Economía de los Datos

1.1 Definición, evolución y situación actual desde una perspectiva económica — 22

2

El *big data* y su importancia en el mundo digital

2.1 Descripción y definición del concepto *big data* — 35
2.2 El componente tecnológico del *big data* — 39
2.3 *Data science*: descubriendo información donde solo había datos — 41
2.4 Economía de los Datos y *big data*, binomio esencial — 43
2.5 Obstáculos y adopción del *big data* en España y América Latina — 46

3

La regulación de los datos en el mundo

3.1 La protección de los datos personales en el mundo — 54
3.2 El enfoque europeo — 55
3.3 América Latina — 60

4

Parte II - Mercados y oportunidades

El mercado de los datos

- 4.1 Agentes — 67
- 4.2 Ciclo de vida del dato: generación, captura, transmisión, almacenamiento, análisis, procesamiento — 72
- 4.3 ¿Dónde se almacena el *big data*? — 80
- 4.4 El análisis de datos y sus profesionales: estado actual, demanda actual, *gap* y previsiones futuras — 81
- 4.5 Evolución del volumen de datos y su utilización en la Economía Digital — 85
- 4.6 *Open data*: concepto, limitaciones y medidas adoptadas en favor de la creación de estándares — 88

5

Nuevas oportunidades de negocio en la Economía de los Datos

- 5.1 La explotación de nuevos modelos de negocio y su encaje en la regulación actual — 99
- 5.2 La intermediación: industrias *data-driven* — 106

6

El auge de nuevas profesiones

- 6.1 Nuevos liderazgos: director de datos - *chief data officer* — 123
- 6.2 Ingeniero de datos - *big data engineer* — 124
- 6.3 Científico de datos - *lead data scientist* — 125
- 6.4 Director de seguridad de la información - *chief information security officer* — 126
- 6.5 Delegado de protección de datos - *data protection officer (DPO)* — 130

7

Parte III - Principales retos de la Economía de los Datos

Titularidad y residencia de los datos — 135

8

Privacidad, soberanía, seguridad y transparencia de los datos

- 8.1 Soluciones a la privacidad y seguridad de los datos personales — 147
- 8.2 Soberanía de los datos: una preocupación global — 155
- 8.3 Seguridad de los datos — 160
- 8.4 Transparencia de los datos — 163

9

La medición de la contribución de la Economía de los Datos en las cuentas nacionales

- 9.1 La metodología de estimación de la Economía de los Datos — 172
- 9.2 La dimensión de la Economía de los Datos — 176

Bibliografía — 179

Prólogo

Explorar y familiarizarse con los elementos básicos de uno, si no el principal, de los vectores de crecimiento del siglo **xxi**, la Economía de los Datos, es un ejercicio que procura facilitar esta publicación.

Todos los agentes estamos llamados a concienciarnos de una realidad que, si bien es relativamente nueva, representa apenas el inicio de una auténtica revolución industrial, sustentada, en esta ocasión, en los datos. Datos que generamos, que almacenamos, que transmitimos, que analizamos, que nos describen, que nos geolocalizan y que desvelan nuestros gustos y preferencias, nuestra opinión e incluso las opiniones de nuestros familiares y amigos. La concienciación a la que aludimos es más necesaria que nunca, tanto para aprovechar al máximo, y en la medida de las posibilidades de cada uno, las bondades aparejadas a las oportunidades que abre la Economía de los Datos, como para protegernos de la exposición que consciente o inconscientemente hacemos de nuestros datos, más o menos sensibles.

El dato se ha convertido en insumo fundamental de cualquier proceso económico. En su estado bruto —sin tratar, aislado—, el dato carece a priori de valor. Es de su tratamiento, procesamiento y análisis científico de donde se extrae conocimiento útil y original —nunca antes se pudieron «leer» los datos como puede hacerse hoy— gracias a las innovaciones en *big data* y en Ciencia de los Datos. Nuevas herramientas (lenguajes de programación, *machine learning*, métodos científicos de datos), nuevas capacidades (Internet en Movilidad e Internet de las Cosas, *cloud computing*) y nuevas destrezas profesionales en el ámbito digital (de las que se da debida cuenta, ya que hoy nos encontramos en un escenario de escasez de perfiles) son las que han hecho posible que en torno al dato se generen modelos de negocio hasta hace poco inimaginables, por inviables.

Extender la frontera del conocimiento y, por ende, de las capacidades innovadoras de las personas, empresas y Administraciones públicas, a partir de la explotación inteligente del dato, es el propósito de la Economía de los Datos, una realidad ya omnipresente en nuestras vidas.

Sobre los autores

Emilio Ontiveros

Fundador y presidente de Afi Analistas Financieros Internacionales y catedrático de Economía de la Empresa de la Universidad Autónoma de Madrid.

Diego Vizcaíno

Socio director del área de Economía Aplicada de Afi Analistas Financieros Internacionales.

Verónica López Sabater y María Romero

Consultoras del área de Economía Aplicada de Afi Analistas Financieros Internacionales.

Alejandro Llorente

Cofundador y *data scientist* de Piplerlab S.L. y profesor en el máster «Data Science & Big Data» de Afi Escuela de Finanzas.

Resumen ejecutivo

Nos encontramos al borde de la denominada cuarta revolución industrial que, como las anteriores, está llamada a transformar buena parte, si no todos, de los ámbitos de nuestra existencia: cómo producimos, cómo consumimos, cómo nos relacionamos y cómo nos comunicamos, en definitiva, cómo evolucionamos como sociedad. La primera revolución industrial estuvo sustentada en innovaciones tecnológicas como la máquina de vapor; la segunda tuvo a la electricidad y el petróleo como detonantes, mientras que la tercera eclosionó gracias al desarrollo de las tecnologías de la información y la comunicación (TIC), que permitieron consolidar el proceso de globalización hoy irreversible. La actual revolución está sustentada en un elemento aparentemente más etéreo, pero, como podrá constatar en la presente publicación, omnipresente: los datos. En la Economía del Conocimiento, los datos son insumo básico y clave.

Sin embargo, a menos que sea objeto de un proceso de refinamiento, procesamiento y análisis que le dote de utilidad, el dato en sí mismo no es generador de valor.

Los datos generados por ciudadanos, empresas, Administraciones públicas y, más recientemente (y solo es el inicio), por los objetos (Internet de las Cosas, IoT) han aumentado exponencialmente en los últimos años y, a la espera del despliegue del 5G, la tendencia es imparable. Es de este crecimiento exponencial del volumen, la variedad y velocidad de generación de datos —de ahí el término *big data*— y de las mayores capacidades para su captura, almacenamiento, procesamiento y análisis (aplicando métodos de la ciencia de datos o *data science*), del que surge la denominada Economía de los Datos. El desarrollo de otros avances tecnológicos tales como el Internet 2.0 (redes sociales) y sucesivas, así como el sistema *cloud* (centros de almacenamiento y procesamiento remotos de datos), han contribuido a reducir las barreras de entrada de empresas y Administraciones públicas a la inversión y el uso de herramientas digitales para optimizar y proteger los datos (germen de la información y del conocimiento).

La Economía de los Datos no puede entenderse de forma aislada de las ya universalmente adoptadas TIC (*smartphone*, fundamentalmente), el desarrollo de la

biometría y la consideración de la economía, la psicología y la sociología, entre otras disciplinas.

¿Qué tipo de preguntas podemos responder con tal cantidad de datos? En términos de generación de valor, el *big data* puede desarrollar un mayor conocimiento sobre el perfil del cliente, una reducción en los costes mediante la detección y la resolución de ineficiencias y la creación de nuevos productos y servicios de información o la implantación de nuevos negocios. Por otro lado, la *data science*, que engloba métodos computacionales, matemáticos y estadísticos, nos permite responder a preguntas dentro del marco empresarial aplicando el método científico, pasando de la teoría a la práctica, describiendo, anticipando y prediciendo sucesos y prescribiendo acciones.

En cuanto a la regulación de los datos, la misma se basa en la salvaguarda de la privacidad de información sensible, generalmente mediante la prohibición de la divulgación o el uso indebido de información sobre personas físicas. La recopilación y el almacenamiento de datos personales por parte de entidades públicas y privadas demandaron una serie de reglas de juego ante la exposición masiva de datos por parte de las TIC, especialmente Internet. Con el inminente desarrollo del IoT sustentado en la captura y el envío autónomos de datos por parte de objetos cotidianos conectados a Internet, además de implicaciones en materia de seguridad, también se desvelan nuevas preocupaciones en materia de privacidad y protección de datos, en este caso no necesariamente de carácter personal. En términos geopolíticos, Europa es la región del mundo en que la protección de datos ha alcanzado un nivel más elevado, mientras que en Norteamérica es la privacidad la que ha alcanzado un alto grado de desarrollo. Los principios y derechos básicos para europeos contemplados en el estándar internacional quedan recogidos en el Reglamento General de Protección de Datos (RGPD) de la Unión Europea.

Mercados y oportunidades

Dentro de la cadena de generación de valor del dato podemos encontrar diferentes agentes involucrados: empresas generadoras de datos, empresas tecnológicas, compañías de servicios analíticos, reguladores y entidades del ámbito académico. La posición de cada agente responde, de alguna manera, a un punto de la cadena de valor del dato.

En la generación de datos, las compañías utilizan varias tecnologías que normalmente se encuentran separadas, por lo que la explotación de los datos que generan se convierte en algo nada trivial, incluso cuando en muchísimas ocasiones, grandes y pequeñas empresas por sí mismas no suponen una fuente de datos masivos. A su vez, dependiendo de la pregunta que queramos resolver, los datos disponibles en la compañía pueden darnos una visión demasiado parcial, que será necesario enriquecer con otras fuentes de información de carácter externo.

Por otro lado, la gran masa de datos supone un reto de capacidad para las empresas. Diferentes tecnologías solo permitían acceder a la Economía de la Información a

las grandes empresas con capacidad para hacer fuertes inversiones en capital físico. Sin embargo, sistemas como el *cloud computing* no solo posibilitan acceder a todas las empresas a este input sino que a su vez permiten ajustar el grado de demanda, dotándolo de flexibilización y eficiencia: transición de la economía «brick and mortar» a la Economía de la Información.

La Economía de los Datos está favoreciendo el surgimiento de nuevos modelos de negocio —tanto locales como globales— que están reconfigurando la estructura de muchos mercados y sectores productivos, permitiendo aumentar la eficiencia en la producción y, sobre todo, en la distribución de bienes y servicios, ya sean estos tradicionales o nuevos. Innovación y transparencia están desdibujando las barreras de entrada en ciertos mercados que se basaban en la información, fomentando una reorganización de los negocios tradicionales al aparecer nuevas empresas que impulsan una mayor competencia en el mercado. Un elemento básico y diferencial en el que se apoyan estas nuevas empresas de la Economía de los Datos es la transparencia. Puesto que los mercados funcionan como transmisores de información —a través de los precios y las cantidades de los bienes y servicios que ofrecen y las decisiones que toman los consumidores y oferentes—, su mayor afluencia y transparencia reducen tanto los costes de transacción como los problemas derivados de la asimetría de la información. De esta forma, se introduce una mayor disciplina entre los operadores tradicionales. Estos sistemas corrigen o alinean a través de la reputación los incentivos de la oferta y la demanda, y permiten una asignación más eficiente de los recursos por parte de ambas partes del mercado.

Las plataformas digitales, que desde un punto de vista económico actúan como intermediarios entre oferentes y demandantes de un mercado que, al igual que en el escenario tradicional, ya han transformado un amplio abanico de industrias (las pioneras son aquellas relacionadas con contenidos educativos, informativos y culturales, transporte, alojamiento o compraventa de bienes y servicios minoristas) y seguirán haciéndolo en otros sectores.

Por regla general, las plataformas digitales se benefician de economías de escala, ya que sus actividades de intermediación implican elevados costes fijos. Además, una característica natural de las plataformas es que cuanto mayor es el número de empresas que ofrecen sus bienes o servicios, mayor posibilidad de elección tendrán los consumidores. Es lo que se conoce como economías de red, cuya existencia no deja de constituir un importante desafío para la regulación, puesto que habilita a las plataformas a erigir nuevas barreras de entrada a medida que las redes se multiplican bajo la misma plataforma. Esto significa que la utilidad de los consumidores en un lado del mercado aumenta conforme lo hace el número de consumidores en el otro lado. Este efecto plantea implicaciones importantes tanto para las estrategias de negocio como para las estructuras de los mercados con múltiples lados.

La revolución digital que supone el desarrollo de la Economía de los Datos ha abierto el debate del impacto que ejercerá sobre el empleo. En la medida en que las

máquinas puedan reemplazar parte o la totalidad de algunas tareas desarrolladas actualmente por trabajadores, gracias al procesamiento de información de manera instantánea (*machine learning*) o los nuevos avances en la producción industrial (como la impresora 3D o el vehículo autónomo), resulta inevitable pensar que buena parte de los puestos de trabajo que conocemos hoy desaparecerán o, como mínimo, sufrirán un proceso de transformación. Sin embargo, la revolución digital también abre un importante y variado abanico de oportunidades laborales que es conveniente conocer y aprovechar.

La demanda de perfiles digitales no solo ha crecido de forma exponencial en los últimos años, sino que está llamada a ejercer un papel relevante en la nueva estructura del mercado de trabajo. Principalmente, destacan estas tres figuras: el director de datos o *chief data officer* (CDO), el ingeniero de datos o *big data engineer* y los directores de seguridad y protección de datos. Los rasgos que se deben destacar de este tipo de perfiles laborales son el dominio de múltiples disciplinas (matemática, estadística, ingeniería, informática y negocios) y la adaptación a entornos cambiantes.

Retos

Aunque la Economía Digital ya es el presente, de cara al futuro todavía persiste una serie de retos y, entre los mismos, destacan la titularidad y la residencia de los datos; la privacidad, la soberanía, la seguridad y la transparencia, y la medición de su contribución a la economía.

La Estrategia Europea del Mercado Único Digital tiene el propósito de mejorar el acceso de los consumidores y empresas a los bienes y servicios digitales en toda Europa, y a la creación de derechos para los productores de datos. Esta legislación procura resolver el *trade-off* entre el derecho a la propiedad intelectual, la privacidad y la innovación.

La sociedad todavía no ha asimilado la infinidad de información que produce y revela diariamente, muchas veces escapando a su control (por ejemplo, en el uso descuidado de las redes sociales). Esto provoca que la preocupación relativa a la privacidad de la información sea creciente, por lo que una serie de mecanismos tales como términos y condiciones de uso, avisos de privacidad y la eliminación o anonimización de contenidos, se están implementando recientemente.

La residencia y el almacenamiento de los datos es un factor clave para la soberanía y la construcción de la Economía Digital, por lo que los Gobiernos procuran adaptar la jurisdicción para mejorar su control. Evitar el uso no autorizado y la corrupción de datos es otro de los retos, por lo que la seguridad se basa en la prevención, en evitar pérdidas de datos y en la monitorización de los canales de transición mediante técnicas de encriptación, *tokenización* y gestión de claves, englobándose todas ellas bajo el concepto de ciberdefensa.

Dado que el mismo dato puede circular por la red durante muchísimo tiempo, el RGPD considera de máxima importancia el cumplimiento del derecho de informa-

ción que todo titular de los datos tiene que tener sobre el tratamiento, el uso y el destino que se aplicarán a los datos personales de su titularidad.

¿Y cuánto contribuye la Economía de los Datos a la economía? Realmente no existe un consenso definido en torno a la contribución de la Economía de los Datos al crecimiento. El sistema tradicional de medición a través del Producto Interior Bruto (PIB) se calcula a partir de las operaciones de compraventa llevadas a cabo en el mercado que llevan asociado un determinado precio, es decir, se basa en lo que se conoce como economía de mercado. Sin embargo, este enfoque convencional para la medición de la actividad económica, que ha servido para calcular el valor de los modelos de negocio tradicionales, no necesariamente puede valer para estimar el volumen de negocio que es capaz de generar la Economía de los Datos.

La revolución de los datos ha impregnado prácticamente todas las esferas de la vida cotidiana, y se ha abierto un debate sobre la capacidad de las estadísticas y metodologías empleadas hasta la fecha para captar, precisamente, las nuevas relaciones comerciales características de la Economía Digital, en general, y de la Economía de los Datos en particular. La cuarta revolución industrial supone, por tanto, un replanteamiento de la manera en la que las estadísticas recogen la realidad socioeconómica que emerge con el desarrollo de la Economía de los Datos, obligando a los sistemas de cuentas nacionales a reinventarse.

Parte I

Introducción

1

Introducción: la Economía de los Datos

1.1

Definición, evolución y situación actual desde una perspectiva económica

22

La progresiva incorporación de los avances acumulados en las últimas décadas en materia de tecnología y redes de comunicación —articuladas en torno a Internet— a la producción de bienes y servicios ha contribuido decisivamente a reconfigurar las relaciones económicas y sociales a escala global. De la época analógica a la era digital, se han sucedido notables transformaciones en todos los ámbitos: desde la manera de consumir bienes y servicios —por ejemplo, a través del *eCommerce*— a la manera de producirlos —los servicios financieros o los medios de comunicación son buenos exponentes— o regularlos.

No hay que entender este proceso como un cambio estructural repentino —pocos en la historia pueden caracterizarse así—, sino que es producto de una evolución paulatina, que alterna periodos de mayor intensidad transformadora con otros de menor profusión. Sin embargo, a lo largo de toda esa línea evolutiva, el elemento nuclear que permanece y que posibilita estas transformaciones es la información. En otras épocas fue la aparición de fuentes de energía o el descubrimiento de innovaciones en el transporte. No es de extrañar, por consiguiente, que se aluda a este episodio de la historia en el que nos encontramos como la «era de la información»; desde los años setenta del siglo xx se viene gestando una revolución en la que los datos pasan a ocupar un lugar central para explicar la realidad económica y social y sus transformaciones.

En el ámbito de la economía, la información adquiere una relevancia creciente a la hora de explicar el producto nacional de los países, dado el protagonismo adquirido por las industrias intensivas en datos y la importancia que reviste como *input* fundamental de producción. En este sentido, los datos se convierten en esenciales para el desarrollo de la Economía Digital, en la medida en que habilitan la innovación, la eficiencia de los procesos y la sofisticación de los bienes y servicios que se producen. Y es que una de las características de la época digital es la producción y el almacenamiento de datos e información a bajo coste, lo que retroalimenta el círculo virtuoso de la innovación.

El proceso evolutivo de transformación económica y social antes aludido se encuentra en un momento de pleno apogeo. La tecnología avanza a pasos agigantados —la inteligencia artificial y el Internet de las Cosas son ya realidades que comienzan a penetrar en la sociedad— y, con ella, surgen nuevos retos, como la preocupación sobre la utilidad que se le brinde a la utilización de la ingente cantidad de datos generados de forma automática, a pesar de ser este el punto donde reside la oportunidad de la transformación digital.

Los datos generados por ciudadanos, empresas, Administraciones públicas y, más recientemente, objetos (Internet de las Cosas, IoT) han aumentado exponencialmente. En muchos casos, se trata de información asociada al usuario, esto es, datos personales que permitirían identificar a una persona concreta, por lo que requieren —los datos y las personas que los generan y/o comunican— una protección especial en ámbitos relacionados con la transparencia, el consentimiento, la finalidad, la protección, la seguridad, la conservación, la calidad y la exactitud, aspectos que se irán analizando en los diversos capítulos de esta publicación.

La seguridad digital, por motivos análogos, se convierte en un aspecto clave en la nueva Economía de los Datos.



En el ámbito europeo, el Mercado Único Digital es considerado uno de los ámbitos de progreso más prometedores y de los que más retos plantean. Los avances que conlleva dicho mercado requieren un marco normativo que facilite el desarrollo de la computación en nube, la conectividad de datos móviles sin fronteras y un acceso simplificado a la información y al contenido, preservando la privacidad, los datos personales o la seguridad cibernética, entre otros.

1.1

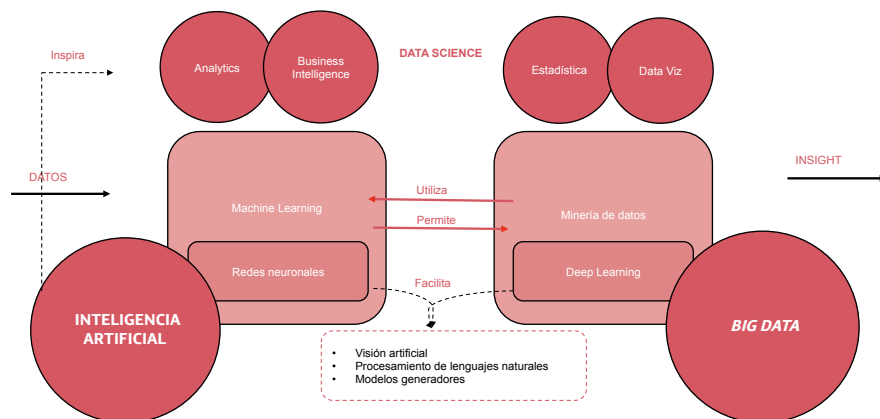
Definición, evolución y situación actual desde una perspectiva económica

1.1.1 Definición: conceptualización y perímetro

En los últimos años, especialmente desde la popularización del acceso y el uso de Internet, los entornos sociales, económicos, culturales y productivos enfrentan cambios que afectan desde las tareas más sencillas —como la comunicación interpersonal o las decisiones de compra cotidiana— a otras más complejas, como el funcionamiento de la cadena de producción de la industria o el *marketing* de una empresa, que están suponiendo una auténtica transformación de carácter digital de corte transversal en todos los ámbitos de nuestra vida. Muchos de estos cambios vienen explicados por la innovación tecnológica, que en el siglo *xxi* ha alcanzado velocidades vertiginosas, habilitada en buena medida por la creciente capacidad para obtener, procesar y analizar datos.

Fig. 1.1

La relación entre *big data*, *data science*, *machine learning*, *deep learning* e inteligencia artificial.



Fuente: whatsthebigdata.com.

La Economía de los Datos puede definirse como el conjunto de iniciativas, actividades y/o proyectos cuyo modelo de negocio se basa en la exploración y explotación de las estructuras de bases de datos existentes (tradicionales y procedentes de nuevas fuentes) para identificar oportunidades de generación de productos y servicios.

El nuevo contexto de relaciones en torno al dato trae consigo nuevas reglas de juego para los agentes —ciudadanos, Gobiernos y empresas— de carácter transversal —no discrimina entre sectores— y supone un cambio en las formas de producción y consumo, de planificación y gestión, de distribución, movilidad, etc., derivadas de los cambios tecnológicos relacionados con la información, las comunicaciones y la globalización.

Los datos son el centro de la nueva economía. Una definición práctica del término «datos» es la que recoge la Norma ISO/IEC 2382-1, que señala que un dato es la «representación reinterpretable de información de una manera formalizada, adecuada para la comunicación, la interpretación o el procesamiento». Dicho de otro modo, los datos son valores que reflejan situaciones o hechos que se producen y que permiten recrear o conocer una realidad. Se trata de una representación simbólica (numérica, alfabética, algorítmica, espacial, etc.) de un atributo o variable cuantitativa o cualitativa que describe hechos empíricos, sucesos y entidades. Un dato puede surgir de distintas fuentes y presentar diversas formas: por ejemplo, una imagen de una cámara de tráfico, una fotografía, un mensaje en las redes sociales, una voz en una conversación, la temperatura de una habitación, el grado de humedad de un cultivo, el número de unidades vendidas por una tienda o la hora exacta en la que se entregó un pedido.

Las imágenes, por ejemplo, están dotadas de metadatos, o información adicional sobre los datos, de contexto, tales como el tipo de extensión del archivo, el tamaño

y la fecha de creación, el dispositivo utilizado, la versión del *software* correspondiente, la ubicación (georreferenciación), la fecha y la hora, etc. Los metadatos son datos que describen otros datos, esto es, ayudan a ubicar datos.

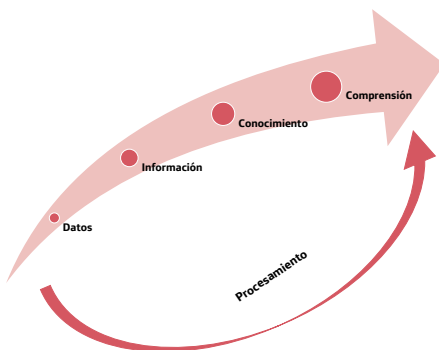
Los datos son el nuevo recurso considerado por muchos «la nueva gasolina» (Kroes, 2011; Rotella, 2012; Arthur, 2013), quienes reflejan su relevancia y la cada vez mayor dependencia de este recurso. También han sido definidos como un recurso infraestructural (OECD, 2014), ya que pueden ser utilizados por un número ilimitado de agentes para un infinito número de aplicaciones para producir bienes y servicios, considerados el equipamiento y estructura básica necesaria para que un país, una región o una organización funcione de forma adecuada (Merriam-Webster, 2004).

Sin embargo, un dato por sí mismo no proporciona valor, igual que la imagen de una cámara de tráfico, sin más contexto o tratamiento, presenta una utilidad muy limitada. Para obtener información de los datos y, por tanto, utilidad de los mismos, es necesario su procesamiento. Por esta razón, una de las metáforas más utilizadas es que los datos, más que gasolina, son el nuevo petróleo (Michael Palmer, Association of National Advertisers), por la necesidad de transformarlos (refinar) para ser convertidos en insumos de valor y utilidad.

Una vez que a los datos brutos (por ejemplo, un sonido) se les agrega información de contexto básica (metadatos), estos se convierten en información (por ejemplo, el sonido pertenece a una máquina) que, por sí misma, no proporciona ningún valor: no se conoce si el sonido es indicativo de una alarma, si se trata de un sonido normal emitido de forma constante por la máquina o si tiene algún otro significado. Por tanto, es necesario considerar un contexto mayor para que el dato pueda ser de alguna utilidad, para que la información se convierta en conocimiento. Por ejemplo, el volumen a partir del cual el sonido que produce la máquina indica un incorrecto funcionamiento de la misma y puede ser un síntoma de fallo de ciertas piezas.

Fig. 1.2

El ciclo del dato



En esta fase, el conocimiento puede ser útil por sí mismo, pues puede aportar información que previamente no era conocida. Sin embargo, en la mayor parte de los casos, es necesario tener una pregunta que responder. Generalmente, los análisis de datos se realizan para dar respuesta a determinadas preguntas: ¿cuál es la ruta óptima para llegar de un punto a otro?, ¿qué probabilidades hay de que llueva mañana?, ¿qué modificaciones en la cadena de producción pueden hacer los productos más rentables?, ¿qué clientes tienen más posibilidades de abandonar la compañía? En el ejemplo anterior, la pregunta podría ser la siguiente: ¿cuál es el sonido que indica que es necesaria una revisión antes de que se rompa la máquina?

El análisis de los datos con las distintas técnicas aplicables que detallaremos en los siguientes capítulos ayuda a dar respuesta a las distintas preguntas planteadas, lo que en sí mismo supone una aportación directa de valor al revelar la respuesta a preguntas que antes, por falta de información o de capacidad tecnológica, quedaban sin responder.

La nueva realidad en torno a los datos determina que mientras que un solo dato (un único sonido) puede ser observado, analizado, comprendido y utilizado, estas cualidades dejan de ser abordables cuando se analizan cientos, miles o millones de datos, a menos que se utilicen nuevas disciplinas y herramientas, entre las que cabe distinguir, fundamentalmente, estas dos:

- *Business intelligence* o inteligencia de negocio: permite explicar el pasado, crear cuadros de mando con visualizaciones de la información o informes de errores.
- *Data science* o ciencia de datos: se basa en las técnicas que, además de realizar un análisis descriptivo, posibilitan anticipar o predecir sucesos futuros (análisis predictivo) e incluso prescriptivos.

La ciencia de datos permite responder preguntas que contribuyen a conseguir mejoras de eficiencia (tiempos de transporte más reducidos, campañas de publicidad individualizadas, diseño de productos optimizado), ahorros de costes (reparaciones de maquinaria en el momento preciso, ofertas exclusivas para aquellos clientes con más riesgo de abandonar la compañía, ahorro de costes energéticos...) y otros muchos tipos de ganancias. Entre estos usos, los más destacados son la capacidad de análisis pormenorizado del cliente, la prevención de fraude, los mecanismos de recomendación, el análisis de redes sociales y el Internet de las Cosas.

Estas nuevas herramientas están siendo utilizadas por distintos agentes, desde empresas tradicionales que se transforman en *data-driven companies* (DDC, por sus siglas en inglés) a *startups* surgidas de forma nativa en la Economía Digital, hasta Administraciones públicas que mejoran sus servicios a partir del aprovechamiento del dato (para diseñar o mejorar sus planes de movilidad o para favorecer la detección del fraude, por poner algunos ejemplos). También existen usuarios particulares que se valen de información pública para crear y/o usar aplicaciones que

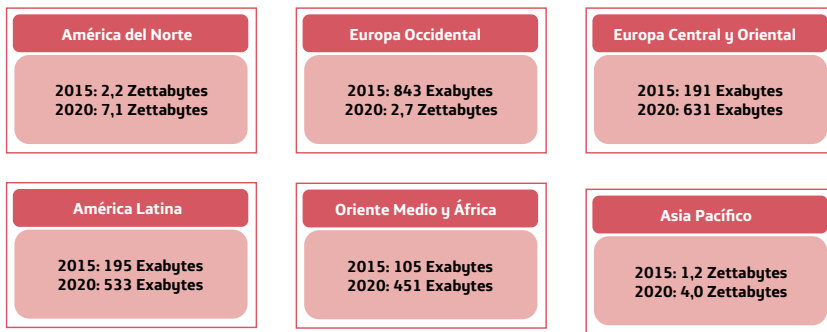
mejoran su día a día (localización de plazas de aparcamiento, análisis de tiempo de llegada de autobuses, servicios *peer-to-peer*, P2P, etc.). No obstante, la adopción de estas nuevas herramientas sustentadas en el dato no es homogénea ni entre sectores ni entre distintos territorios, de manera que sus efectos en algunos casos han quedado más patentes, mientras que en otros aún no han llegado a su máxima capacidad.

1.1.2 Evolución y situación actual desde una perspectiva económica

Una de las claves más destacadas que ha llevado al surgimiento de la Economía de los Datos ha sido precisamente el crecimiento en el volumen de este recurso. Según la Software & Information Industry Association (SIIA, 2013), en 2013 la información almacenada en el mundo alcanzó los 161 exabytes (EB = 10^{18} bytes) por año —5 EB, en 2003—. En 2020 se prevé que el tráfico de datos desde y hacia los centros de datos supere los 15 zettabytes (ZB = 10^{21} bytes).

Fig. 1.3

Tráfico global en centros de datos, por región (2015-2020)



Fuente: CISCO Global Cloud Index 2015-2020.

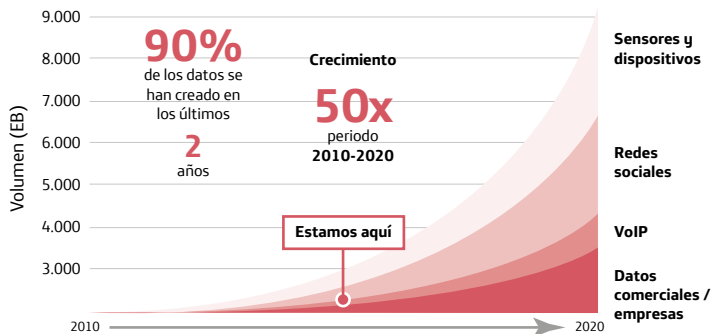
Nota: 1 ZB = 10^3 EB = 10^6 PB = 10^9 TB = 10^{12} GB = 10^{15} MB = 10^{18} kB = 10^{21} bytes.

Según José María Álvarez-Pallete,¹ presidente ejecutivo de Telefónica, «los recursos más valiosos en el futuro serán los datos»; hoy la mitad del tráfico en Internet es de origen no humano. Además, según Telefónica, el volumen de datos en los próximos años se multiplicará por once, la velocidad de transferencia de información será cien veces la disponible actualmente y la capacidad se multiplicará por cien.

1. https://cincodias.elpais.com/cincodias/2017/06/19/companias/1497861452_912952.html

Fig. 1.4

Volumen de big data

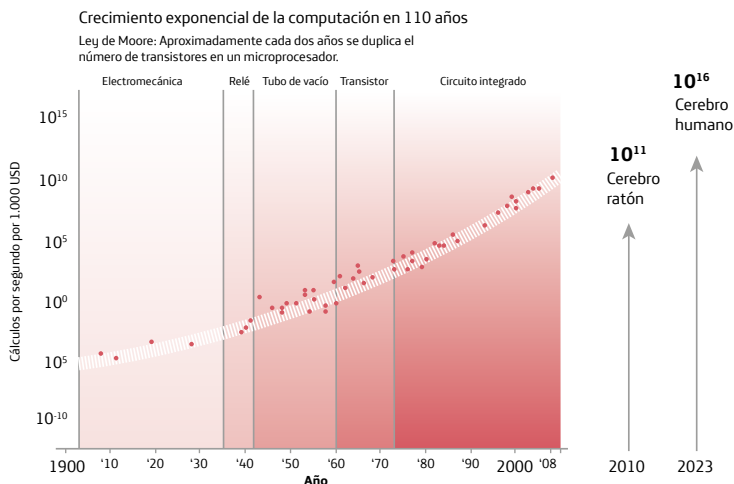


Fuente: (Familian 2016) Working with Big Data.

A pesar de ello, el incremento en el volumen de información no es el único factor relevante que ha permitido esta aceleración. La gran variedad de tipos de datos capturados y analizados, y la propia capacidad de procesarlos para obtener valor, son parte consustancial del elemento revolucionario que ha provocado este cambio global. El núcleo principal de estos avances ha estado en el abaratamiento de la computación, exponencial en los últimos años, con el que la expansión de la Economía de los Datos se ha reforzado.

Fig. 1.5

Evolución de la velocidad de cálculo



Fuente: Kurzweil 2005.

Este entorno favorecedor se ha visto además enriquecido por la convergencia de otros avances como el crecimiento de los objetos conectados (IoT) y en movilidad, las redes sociales, la computación en la nube (*cloud computing*) y el despliegue de herramientas analíticas. El futuro que se avecina con las capacidades esperadas de la computación cuántica supondrá con seguridad una revolución en el procesamiento de los datos y, por ende, en la economía con ellos generada.

El *cloud computing* o procesamiento en la nube permite utilizar capacidad de computación remota, masiva y escalable. Está basada en la servificación, es decir, la conversión de la capacidad de procesamiento computacional de un producto a un servicio contratable por una tarifa que posibilita hacer uso de unos recursos mientras están siendo utilizados. Este uso bajo demanda de la computación evita la necesidad de realizar una gran inversión en infraestructuras, plataformas y *software* y se encuentra a disposición del usuario de forma más directa. Gracias a este servicio, el análisis de datos deja de estar únicamente a disposición de grandes empresas con una gran capacidad de inversión, y pasa a estar al alcance de todos; se nivela el terreno de juego para empresas y Administraciones públicas de cualquier tamaño, al eliminar parte de las barreras de entrada.

Otro de los elementos clave ha sido la proliferación de los dispositivos conectados a Internet, entre los que se encuentran los teléfonos móviles y las tabletas como elementos principales, pero a los que se suman sensores, relojes inteligentes, coches o incluso hogares conectados. Todos estos dispositivos ofrecen nuevas oportunidades gracias a la conectividad prácticamente ininterrumpida, el acceso a la computación en la nube desde casi cualquier localización y la capacidad de acceso a cualquier fuente de información online o la conexión a cualquier dispositivo en cualquier momento. El elemento más relevante, si cabe, al que dan acceso estos dispositivos conectados es la creación de nuevos servicios personalizados basados en los datos recabados de forma autónoma, tales como recomendaciones de salud según las constantes vitales, alertas de retraso en medios de transporte según localización, ofertas de restaurantes o tiendas basadas en los gustos, o alertas de posible fraude ante la detección de situaciones anómalas. Servicios, en definitiva, que permiten ajustarse al máximo a las preferencias y los estilos de vida de los clientes o usuarios.

En relación con la ubicuidad de Internet, el tercer elemento que favorece el gran desarrollo actual es la interconectividad social. Las redes sociales son un fenómeno que ha visto una expansión global en los últimos años, no solo en número sino también en capacidad, incorporando componentes de otros programas, nuevos servicios y diversas plataformas, consiguiendo una conjunción de funciones que alcanzan desde la capacidad de compartir historias, imágenes o vídeos hasta capacidades relacionadas con el mundo laboral o la formación. El análisis de la información recabada en las redes sociales ha permitido un incremento en la comprensión de los individuos y una mejora en los servicios ofertados por distintos proveedores de forma personalizada, tanto dentro de la comunicación, como del mercado de trabajo y el entretenimiento, entre otros muchos.

Al mismo tiempo, las capacidades tecnológicas surgidas en relación con estos puntos han permitido la aparición de nuevos modelos de negocio basados en el entorno colaborativo, ya sea por la posibilidad del libre acceso y reutilización de datos (*open data*), como por las numerosas plataformas y aplicaciones habilitantes de nuevos modelos de negocio (*apificación*) sustentados en la compartición o intercambio de activos P2P (vivienda, vehículos, objetos, tiempo, etc.), la información a usuarios sobre el estado de la movilidad en las ciudades y carreteras (atacos, accidentes o retrasos en medios de transporte) o la puesta en común de ideas para el desarrollo de soluciones y/o iniciativas emprendedoras.

Por último, los avances en capacidad de computación, almacenamiento, sensores y tecnologías de la comunicación han sido el factor determinante en el avance del análisis de datos. Las empresas tienen ahora a su alcance programas estadísticos, de análisis predictivo, prescriptivo, de visualización y de aplicación para la toma de decisiones, independientemente del lugar donde se encuentren los datos, de su volumen, heterogeneidad, etc. Estas herramientas de análisis permiten a los agentes analizar y utilizar su información y la generada por terceros, ya sean datos estructurados o no estructurados, en tiempo real o estáticos, para innovar y mejorar los procesos, productos y servicios ofrecidos.

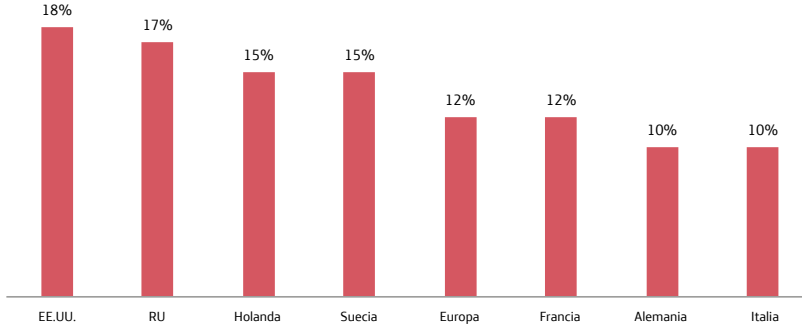
Aunque las tecnologías mencionadas son las más relevantes, hay muchas otras que se engloban en esta corriente y que han evolucionado a lo largo de los años. Prueba de esta vertiginosa evolución tecnológica es la diferencia existente entre las sucesivas curvas de expectativas tecnológicas elaboradas por Gartner en su *Hype Cycle Report* desde 2005.

Es aún difícil cuantificar el impacto económico de la Economía de los Datos, debido a su amplia diversificación, su gran alcance y su alta capacidad de transformación. Según la Comisión Europea, se estima que la Economía de los Datos en la Unión Europea ascendía a un valor de 272.000 millones de euros en 2015. McKinsey (Henke, et al. 2016), por su parte, estima que el PIB de la economía europea podría incrementarse en 2,5 billones en 2025 si alcanzara un mayor potencial, especialmente en las empresas más retrasadas en términos de digitalización. En el caso de Estados Unidos, McKinsey estima que el Internet de las Cosas, la analítica de datos y las plataformas de talento online podrían añadir 2,2 billones de dólares al PIB de la economía americana en 2025.

Sin embargo, a pesar de que el 90% de los datos digitales ha sido generado en los últimos dos años, solamente un 1% de esa información ha sido analizada. Se estima que Europa ha alcanzado apenas un 12% de su potencial, mientras que Estados Unidos se encuentra en el 18%.

Fig. 1.6

Grado de aprovechamiento del potencial de la digitalización en varios países



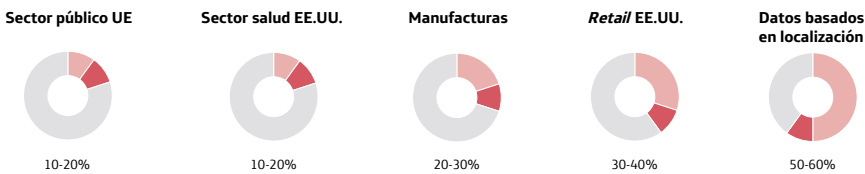
Fuente: McKinsey Global Institute Analysis.

Nota: Europa representa el promedio de los seis países mostrados, que suponen el 60% de la población, el 72% del PIB de la UE-28.

La mayor parte de las compañías captan aún hoy una mínima parte del valor potencial de sus datos. De entre los dominios existentes, el mayor progreso se ha producido en el área de servicios basados en localización y las ventas al por menor, gracias a los negocios nativos digitales que han surgido en ambos ámbitos. En el sector manufacturero, las Administraciones públicas y la salud, se ha conseguido capturar menos del 30% del valor potencial previsto hace cinco años.

Fig. 1.7

% del valor capturado (2016) respecto al valor previsto en 2011



Fuente: (Henke, et al. 2016) <http://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/the-age-of-analytics-competing-in-a-data-driven-world>.

2

El *big data* y su importancia en el mundo digital

2.1

Descripción y definición del concepto *big data*

35

2.5

Obstáculos y adopción del *big data* en España y América Latina

46

2.2

El componente tecnológico del *big data*

39

2.3

***Data science*: descubriendo información donde solo había datos**

41

2.4

Economía de los Datos y *big data*, binomio esencial

43

Al *big data* frecuentemente se le caracteriza mediante tres uves: volumen, variedad y velocidad.

- Volumen es el atributo más obvio recogido en el propio término de *big data*. No en vano se ha observado una evolución desde magnitudes como los megabytes, gigabytes o terabytes hacia los petabytes.
- Variedad tanto en tipología de datos como sus fuentes, de forma tal que se pasa de manejar datos estructurados en bases de datos (procedentes de fuentes limitadas) y estáticos o cuasi estáticos a tratar (i) datos estructurados, semiestructurados y desestructurados; (ii) datos dinámicos o en continuo cambio, y (iii) originados por personas, máquinas, sensores, etc.
- Velocidad en la captura, el movimiento y el proceso de los datos, llegando a ser en tiempo real.

Algunas organizaciones han añadido nuevas uves a la definición del *big data*: veracidad (la calidad de los datos capturados), variabilidad (manejo de inconsistencias por el cambio en el significado de los datos) y valor (los ingresos o beneficios derivados).

2.1

Descripción y definición del concepto *big data*

Con afirmaciones y símiles tan llamativos como «el petróleo del siglo XXI»,¹ el *big data* se ha posicionado como una de las grandes revoluciones tecnológicas que están sucediendo hoy en día. Este término hace referencia a la aparición y el aprovechamiento de grandes volúmenes de datos que, si bien algunos de ellos existían antes, nunca hasta ahora se habían contemplado como una fuente de valor para las empresas, de transparencia para las Administraciones públicas o de mejor autogestión para todos los ciudadanos en su día a día. De la mano de la digitalización —transformación por la cual procesos que anteriormente se realizaban de una manera física, manual o mecánica, ahora se realizan mediante tecnología informática—, todo lo que sucede en nuestra vida cotidiana deja un rastro en forma de dato.

¿Y qué es un dato? Un dato es un registro que se almacena en silos y lagos de información, y se guardan variables como el actor que produjo una interacción, en qué momento se produjo, en qué lugar, qué características tenía dicho actor, en qué contexto sucedió, etc.

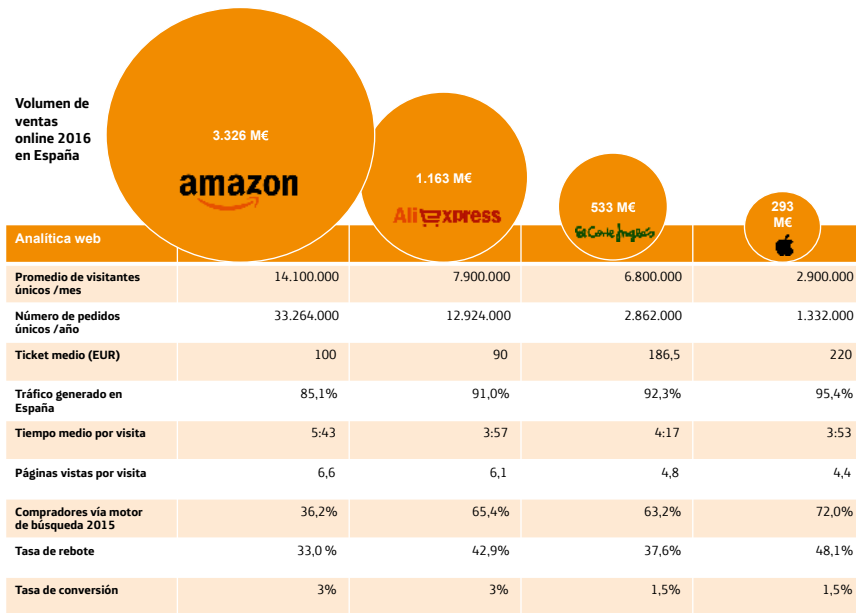
Por poner un ejemplo cotidiano, cada vez que un usuario realiza una compra a través de un comercio online, se establece una huella en alguna base de datos de

1. «The world's most valuable resource is no longer oil, but data», disponible en: <https://www.economist.com/news/leaders/21721656-data-economy-demands-new-approach-antitrust-rules-worlds-most-valuable-resource> (*The Economist* 2017).

qué compró, cuánto pagó, cómo pagó, qué productos compró a la vez, dónde estaba cuando realizó la compra... y, así, un sinfín de información. Pensemos ahora que, solamente en España, los grandes portales de comercio online están facturando cifras crecientes cada año, involucrando miles de millones de transacciones y millones de usuarios con toda la información mencionada anteriormente. Por eso es *big*.

Fig. 2.1

Los gigantes del comercio en la web



Fuente: <http://www.expansion.com/economia-digital/companias/2017/08/20/59970402e5fdeadf3e8b460a.html>

La digitalización no solamente ha permitido sensorizar muchos de nuestros movimientos y acciones, sino que ha facilitado escalar la economía de manera nunca vista anteriormente. Este proceso ha favorecido que el comercio trascienda gran parte de las barreras físicas que tradicionalmente tenía: una zapatería tenía un límite de negocio que venía dado por el tamaño de la tienda, el número de dependientes o el horario de apertura; a día de hoy, cualquier empresa textil con un portal online puede dar servicio en cualquier punto del mundo y los clientes pueden adquirir sus productos a cualquier hora en cualquier lugar. Por supuesto, siguen existiendo límites: los productos físicos siguen dependiendo de una logística, de unos proveedores y de una fabricación que es difícil que se escalen al mismo ritmo que las posibles ventas, pero la propia digitalización de la industria y su monitorización continua per-



miten optimizar los procesos hasta límites que aún no somos capaces de vislumbrar. Lo que ha surgido como otro de los términos de moda, la transformación digital, tiene como consecuencia natural la generación de volúmenes de información anteriormente nunca vistos.

2.1.1 ¿Dónde está el *big data*? Grandes cifras y fuentes de datos

No solo tenemos disponibles los datos que ahora somos capaces de sensorizar a través de la transformación digital. Existen terceras partes que permiten utilizar sus datos, respetando los aspectos legales asociados, para construir información o aplicaciones sobre ellos.

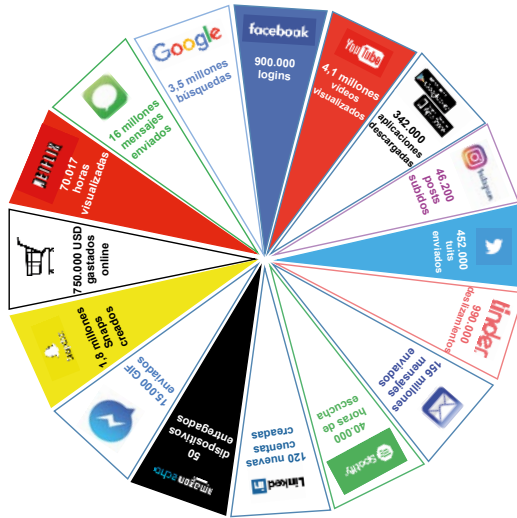
Es el caso de las redes sociales como Twitter o Facebook, que proveen de API (interfaces de programación de aplicaciones o, de forma más coloquial, puntos de conexión de descarga de datos) a terceros, lo que permite obtener información detallada sobre los perfiles de comportamiento y los usuarios de las mismas. YouTube tiene otras API para extraer estadísticas de los vídeos que alberga su portal, y de la misma manera sucede en sitios como LinkedIn, Instagram o Google Plus.

Este *big data* que estamos generando los usuarios externamente a empresas como entidades financieras, aseguradoras, grandes distribuidores, supermercados o marcas de moda son aprovechadas por estas para conocer mejor nuestros gustos e intereses, adecuar sus productos y ofertas o entender cómo son nuestros patrones de movilidad diarios, y adaptar sus campañas de *marketing* al momento en que menos nos molesten. Así, las bases de datos internas de las empresas ya se pueden considerar *big data*: el volumen de datos que se manejan en estas fuentes externas es increíble.

Como se puede observar en la figura 2.2, en solamente un minuto, casi medio millón de tuits son enviados, se realizan tres millones y medio de búsquedas en Google o casi un millón de usuarios se identifican para entrar en Facebook. Cada uno de estos hechos es una fila nueva en una tabla inmensa de datos que refleja cada día lo que hacemos, lo que sentimos o dónde estamos.

Fig. 2.2

Qué ocurre en Internet en un minuto



Fuente: Afi, basado en la creación de @LoriLewis y @OfficiallyChadd. <http://www.visualcapitalist.com/happens-internet-minute-2017>.

¿Acaso antes de esta revolución digital no existían los datos? Por supuesto que sí. Entonces, ¿es esto del *big data* y el análisis de datos algo completamente nuevo? Por supuesto que no.

Uno de los ejemplos más antiguos que se recuerdan de recopilación de la información y explotación de la misma es el caso del análisis de las muertes por cólera realizada por el doctor John Snow, uno de los precursores de la epidemiología, en el Londres de 1854.² Mediante la visualización geográfica de las muertes que se habían producido por cólera en el mes de septiembre de ese año, este doctor verificó que la variable que mejor las explicaba era la existencia de un pozo de agua contaminado en Broad Street. Gracias al análisis de datos, también explicó por qué enfermó gente que vivía lejos de esta calle o por qué de un taller cercano con más de quinientas personas solamente enfermaron cinco empleados.

Los motivos por los que, primero, se ha acuñado un nuevo término para hacer referencia a estas tecnologías y técnicas y, segundo, de repente, se ha convertido en tendencia en todos los sectores económicos, son muy variados.

El principal es el anteriormente mencionado: la transformación digital ha hecho que se haya empezado a recopilar información a gran escala de diferentes procesos empresariales.

2. Entrada en la Wikipedia sobre el brote de cólera en Londres en el año 1854: https://en.wikipedia.org/wiki/1854_Broad_Street_cholera_outbreak

El segundo motivo es la aparición de tecnologías que nos permiten almacenar y procesar volúmenes de información de una manera barata y escalable.

El tercero es la adquisición de conocimiento en el mundo empresarial de que existen una gran variedad de técnicas estadísticas, matemáticas e informáticas, tradicionalmente conocidas desde los años cincuenta en el sector académico y científico, que nos permiten descubrir patrones en nuestros datos y, con ellos, anticipar lo que ocurrirá en el futuro.

El cuarto y último motivo es, precisamente, lo que se ha dado en llamar Economía de los Datos, la consciencia de que tanto los datos en sí mismos como lo que podemos derivar de ellos (información, conclusiones, predicciones, etc.), pueden ayudar a las empresas a generar más ventas, disminuir costes e, incluso, generar nuevos negocios por sí mismos.

2.2

El componente tecnológico del *big data*

La aparición de todas las fuentes de datos mencionadas anteriormente, entre otras muchas, ha supuesto un reto tecnológico a las compañías, ya que los sistemas de información de los que disponían no estaban preparados para hacer frente a las nuevas características que estas tienen en comparación con las fuentes de datos tradicionales.

Los retos principales que surgieron tras la aparición de las fuentes de datos digitales fueron los siguientes:

- Almacenamiento de enormes volúmenes de datos: seguramente, el reto más evidente es la ingente cantidad de datos que era necesario almacenar y procesar.
- Ingesta de datos de múltiples fuentes: aparecen diferentes puntos de acceso a datos, con formas distintas de conexión, formatos, etc.

En los análisis avanzados que las técnicas de aprendizaje automático nos permiten hacer, es necesario modelizar nuestro problema de la manera más rica posible, por lo que es necesario incorporar fuentes diversas de información, tanto internas de la compañía (herramientas actuales, operacionales distintas, aplicaciones de *marketing*, etc.) como externas (redes sociales, datos públicos, meteorología, eventos, localizaciones, etc.).

- Tasas de captura de información: algunas de estas fuentes no solamente generan un volumen muy grande, sino que lo hacen a velocidades desiguales a lo largo del tiempo, con picos enormes.

A modo de ilustración, aunque la tasa de tuits por minuto que hacen mención a un jugador de fútbol sea alta, cuando este mete un gol, en un espacio muy corto de tiempo se produce una cantidad enorme de menciones.

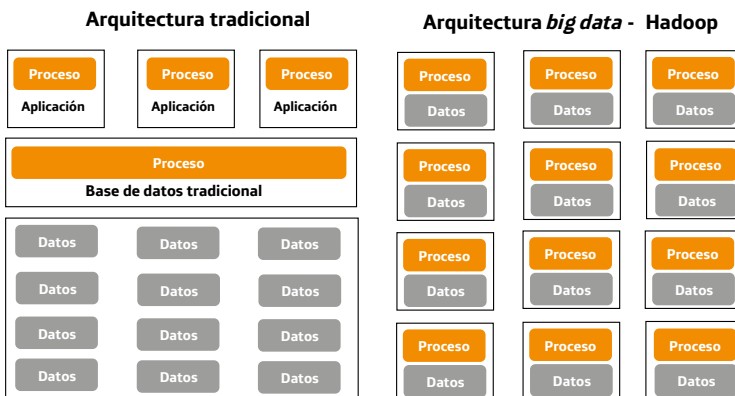
- Datos desestructurados: aparecen fuentes de datos que, en vez de aportar información concreta semánticamente, es necesario preprocesar para extraer su verdadero significado.

Por ejemplo, en una base de datos de clientes de una compañía, existe información como la edad o la ciudad donde vive (campos con información única semánticamente), pero también se recogen las opiniones que dichos clientes escriben en foros, en texto libre y, por tanto, no por el hecho de almacenarlo las máquinas son capaces de entender su significado (¿qué usuarios se han quejado de mi servicio técnico en el último mes?).

Las bases de datos tradicionales estándares, denominadas relacionales, eran muy robustas desde el punto de vista de los procesos y operaciones de las compañías, y aseguraban su consistencia, durabilidad y aislamiento a lo largo del tiempo, pero no eran suficientemente eficaces a la hora de tratar con los problemas mencionados anteriormente. Para solucionarlo, grandes compañías, como fueron los casos de Google y Yahoo, hicieron muchos esfuerzos invirtiendo en investigación y desarrollo, y el resultado fue un cambio absoluto de paradigma en lo que a los sistemas de almacenamiento y procesamiento de la información se refiere. Sin entrar en detalles técnicos, dieron con una solución por la cual se puede almacenar y procesar la información de manera distribuida (entre muchos servidores) con unos requisitos de estructuración de datos mucho menores que los sistemas tradicionales. Esto permite de forma muy sencilla lo que se conoce como escalabilidad horizontal, la competencia de hacer crecer la capacidad de almacenamiento y procesamiento a lo largo del tiempo simplemente agregando nuevos servidores a nuestra infraestructura sin impacto en lo que existía anteriormente.

Fig. 2.3

Escalabilidad horizontal



Fuente: (White 2015) *Hadoop, The Definitive Guide*, O'Reilly Media, 2015.

La principal tecnología que surgió, en 2008, fue Hadoop, una implementación del esquema distribuido previamente mencionado. Pero el principal avance de la aparición de esta tecnología no es solamente el hecho de que resolviera los problemas que las nuevas fuentes de datos proponían, sino que Yahoo y Nutch (las dos principales empresas involucradas en su desarrollo) lo liberaron como un proyecto *open source*, lo que hizo que una gran comunidad de desarrolladores ayudara a que evolucionara, y se estableciera en las compañías tecnológicas como un estándar. Posteriormente han aparecido un sinnúmero de tecnologías en el entorno de Hadoop, y hoy en día existen otros modelos de procesamiento distribuido que lo superan en rendimiento. Pero fue la apuesta estratégica de compartir de manera abierta este conocimiento lo que hizo que el *big data*, en su parte tecnológica, se haya convertido en uno de los principales objetivos de las compañías.

2.3

Data science: descubriendo información donde solo había datos

Las tecnologías mencionadas ayudan diariamente a las compañías a gestionar los datos que producen y que captan de otras fuentes, pero el hecho de poder almacenarlas y procesarlas no implica que extraigamos valor de ellas.

Imagínese una inmensa tabla de datos en la que usted puede buscar, agregar o transformar información, pero ¿para qué? ¿De qué les serviría a las compañías e instituciones públicas tener estos grandes centros de datos si no supieran qué hacer con ellos?

Aunque originariamente el término *big data* hiciera referencia a tecnología, se ha hecho tan popular que muchas veces se utiliza también para hablar de las aplicaciones que tienen los datos. Típicamente, en cualquier encuesta sobre cómo utilizan el *big data* las empresas, nos encontraremos con casos de uso y formas de extraer valor de los datos, más allá de la tecnología que se use.

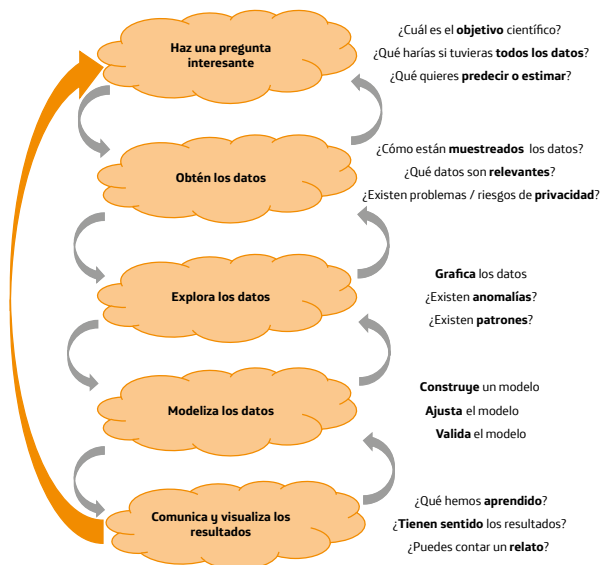
Uno de los principales retos que tienen las organizaciones a la hora de abordar sus proyectos *big data* es precisamente definir qué se quiere solucionar con los datos. Esta tarea requiere, por un lado, el suficiente conocimiento de negocio como para establecer los requerimientos de la organización y estimar la ganancia que se va a obtener tras la realización de estos proyectos, pero también exige, por otro lado, el conocimiento técnico de saber transformar un problema de negocio o un caso de uso en un problema científico viable donde, basándonos en la evidencia que los datos nos dan y utilizando herramientas analíticas, estadísticas y matemáticas, extraer conclusiones sólidas, proponer acciones basadas en datos o predecir comportamientos futuros. Este otro tipo de tareas, con un componente de conocimiento del negocio y ciencia, es lo que se conoce hoy en día como ciencia de datos (*data science*) y que tiene una naturaleza completamente diferente al problema tecnológico previamente tratado.

Los científicos de datos, uno de los perfiles profesionales más demandados hoy en día en el mercado laboral, son precisamente los encargados de llevar a cabo estas funciones. Con un conocimiento que mezcla aptitudes matemáticas, estadísticas, informáticas y de negocio, ellos son los responsables de abordar proyectos y desarrollos que tienen partes muy diferenciadas:

- Identificación del problema: planteamiento de la pregunta que, en caso de ser contestada, resuelve un problema concreto de negocio.
- Obtención de la información: estudio de qué fuentes de datos son las relevantes para solucionar el problema.
- Exploración y análisis de la información: mediante técnicas de análisis estadístico y de visualización, búsqueda de respuestas desde un punto de vista descriptivo al problema concreto.
- Modelización de los datos: seguramente, otra de las principales disrupciones en el mundo *big data*. La posibilidad de crear modelos matemáticos basados en millones de registros históricos que permitan, entre otras cosas, descubrir patrones de comportamiento, predecir cómo actuarán nuestros clientes y usuarios, desarrollar sistemas de recomendación automáticos, etc.
- Comunicación de los resultados: todas estas tareas técnicas y los resultados que se derivan de las mismas, los científicos de datos deben traducirlas al lenguaje del negocio de manera que sea comprensible por los decisores de las organizaciones.

Fig. 2.4

El proceso de la ciencia de datos



Fuente: Bitzstein y Hanspeter (2015) creado para el curso «Data Science» de Harvard <http://os109.org>

En definitiva, la *data science* es el conjunto de técnicas que nos permiten pasar de un almacén de datos a las aplicaciones de valor con los mismos. Aunque no lo veamos, estos modelos ya están presentes en nuestras vidas: cuando encontramos un navegador y estima el tiempo a nuestro destino basándose en el pasado y en datos de tráfico, existe un modelo matemático por detrás; cuando entramos a un portal de *eCommerce* y, basándose en lo que hemos comprado y nuestro perfil, nos recomienda productos de acuerdo a nuestros gustos, es un sistema de recomendación automático el que elige las propuestas; cuando una compañía decide abrir una nueva franquicia en un lugar y no en otro, hay todo un estudio basado en datos de los factores de éxito que influyen, como el tipo de personas que pasan por delante, los flujos de movilidad o el perfil socioeconómico de la zona.

Estamos rodeados de algoritmos y modelos que nos facilitan la toma de decisiones, y que nos conocen gracias a la huella de datos que dejamos, tanto en el mundo virtual como en el físico.

2.4

Economía de los Datos y *big data*, binomio esencial

El *big data* en la actualidad / Monetización de los datos: interna, externa, nuevos modelos de negocio, *open innovation*, productos basados en datos, servicios de información.

En este punto, ya sabemos cómo las empresas son capaces de abordar el problema tecnológico que supone tratar con los volúmenes de información apabullantes que existen hoy en día, y cómo, una vez resueltos los problemas tecnológicos, abordan proyectos con ellos, pero ¿cómo se obtiene valor real de los mismos y qué tipo de proyectos se abordan según los distintos sectores de actividad?

Los proyectos de *data science* que se abordan en las empresas pueden tener diferentes objetivos, pero los principales son los siguientes:

- Conocimiento de los clientes y usuarios: analizando los datos de comportamiento de los clientes, podemos dirigir estrategias, aumentar las ventas y fidelizar dichos clientes manteniendo una mejor relación con ellos derivada de este conocimiento.

Dependiendo del negocio de las compañías, esto toma diversas formas: si estamos en un negocio digital, estudiar y tomar decisiones en función de la navegación de los usuarios, los contenidos que visitan o los factores que influyen en la usabilidad de nuestros portales, puede hacer que nuestras conversiones se multipliquen; si nuestra empresa dispone de tiendas físicas, y somos capaces de medir los lugares por los que pasa un cliente y relacionarlo

con lo que finalmente acabe comprando, podremos cambiar la estrategia y, con ello, aumentar las ventas.

- Reducción de costes: los datos internos de las empresas no son más que un reflejo de lo que está sucediendo y permiten descubrir ineficiencias en los procesos y jerarquías organizacionales, que corregir a partir de los análisis. Un caso tradicional que ayuda a optimizar los costes es la predicción de la demanda que va a tener una empresa en diferentes plazos de tiempo. Por ejemplo, en el sector industrial, para poder ramificar sus servicios de una manera amplia a lo largo del territorio, las empresas cuentan con *partners* en mantenimiento y logística, cuyos acuerdos se regulan de forma variable en función de la demanda prevista. Si la compañía conoce de manera anticipada la demanda que va a tener, podrá ajustar estos servicios, ahorrando costes innecesarios y, además, aumentando el grado de cumplimiento del servicio que otorga a sus clientes.

- Creación de nuevos productos y servicios de información: la propia huella que los usuarios dejan en las bases de datos de las empresas puede permitirles dar recomendaciones de comportamiento a sus clientes con el objetivo de optimizar sus gastos o mejorar sus condiciones.

Un claro ejemplo de esta nueva forma de extraer valor de los datos son los portales inmobiliarios online. Estos portales permiten a los usuarios encontrar ofertas de pisos y locales comerciales, ya sea para alquiler o venta, y ponerse en contacto con los propietarios. Estas compañías disponen, por un lado, de precios de inmuebles repartidos por todo el territorio y, por otro, de los patrones de navegación de los usuarios, que son una muy buena aproximación de la demanda de zonas. Juntando toda esta información, pueden crear herramientas de tasación automática de viviendas e, incluso, ofrecer servicios de información a terceros como una predicción del precio de los inmuebles según sus características y la zona geográfica donde están ubicados.

- Nuevos negocios: los datos de las empresas son útiles para terceras empresas con negocios completamente diferentes, por lo que pueden pensar en crear líneas totalmente nuevas que sean una fuente de ingresos basadas en los datos.

Un caso de nuevo negocio para el sector bancario puede ser crear un producto *software* basado en los datos que tiene de transacciones presenciales con tarjetas de crédito donde, por ejemplo, las Administraciones públicas puedan entender cuál es el comportamiento de los turistas en las ciudades, en qué zonas gastan más según su procedencia, o qué patrones de visitas a puntos de interés siguen.

Fig. 2.5

Mapa de calor de los problemas de negocio *big data* por tipo de industria

	Manufacturas y recursos naturales	Medios de comunicación	Servicios	Gobierno	Educación	Retail	Banca	Seguros	Salud	Transporte	Servicios públicos
Mejorar la experiencia de usuario	52%	78%	66%	43%	76%	83%	77%	77%	73%	69%	44%
Eficiencia de procesos	45%	33%	35%	49%	65%	43%	41%	50%	73%	69%	78%
Marketing	43%	89%	53%	17%	41%	78%	66%	58%		38%	17%
Reducción de costes	42%	33%	35%	37%	35%	30%	41%	31%	45%	56%	61%
Mejora en gestión de riesgos	14%	22%	29%	29%	35%	22%	52%	58%	55%	31%	61%
Nuevos productos	23%	67%	37%	14%	24%	35%	27%	50%		19%	33%
Desarrollo de información de productos	26%	33%	44%	31%	12%	22%	23%	19%	9%	19%	11%
Mejora de las capacidades de seguridad	17%	22%	21%	34%	29%	13%	27%	27%	9%	19%	28%
Cumplimiento normativo	11%	22%	18%	23%	18%	9%	25%	23%	27%	31%	44%

Fuente: Gartner, «State of big data adoption», 2015, disponible en www.slideshare.net/denisreimer/big-data-industry-insights-2015.

¿Y todo ello ya es una realidad? ¿Están consiguiendo las empresas sacar partido de sus datos en estas cuatro vertientes?

Una compañía que conozca mejor a sus clientes será capaz de fidelizarlos más, de aumentar su satisfacción, y de que adquieran un mayor número de productos, por lo que, en general, tendrán un mejor rendimiento comercial. Un caso de éxito muy conocido es el de la cadena de hoteles Four Seasons: dicha cadena analizó el rendimiento de sus restaurantes comparativamente, e ideó una experiencia ideal que todo cliente debería vivir en sus visitas. Con ello, consiguieron aumentar el tique medio de los restaurantes y la satisfacción general de sus clientes. Además, adaptaron los turnos de reserva según la época del año de manera que, en periodos vacacionales, los tiempos eran más largos, puesto que sus clientes no tienen prisa y, por tanto, consiguieron reducir los tiempos de espera.

Si con el *big data* se puede conocer mejor los procesos internos de las empresas, se encontrarán ineficiencias que puedan solucionarse y, por tanto, reducir los costes. Un ejemplo de estas aplicaciones es el caso abordado por Intel en 2012, cuando consiguió abaratar su tecnología gracias al estudio de su enorme histórico de producción. Lo que hicieron fue analizar el proceso de control de calidad, y focalizaron en qué tests eran los más útiles dependiendo de los materiales y componentes de los productos. Si, de media, un componente pasaba por unos diecinueve mil tests, tras este estudio, focalizaron en solamente unos pocos componentes, sin reducir el tiempo de vida de los productos, lo que supone una reducción de un coste directo e inmediato para la compañía.

Finalmente, aunque seguramente es la dimensión más visionaria de la monetización del dato, también existen negocios basados en los datos de clientes a través de servicios de información que nos ayuden a tomar mejores decisiones, aunque esos datos no pertenezcan a la compañía. Un buen ejemplo son las aplicaciones para gestión de las finanzas personales. Aplicaciones como Mint (muy utilizada en Estados

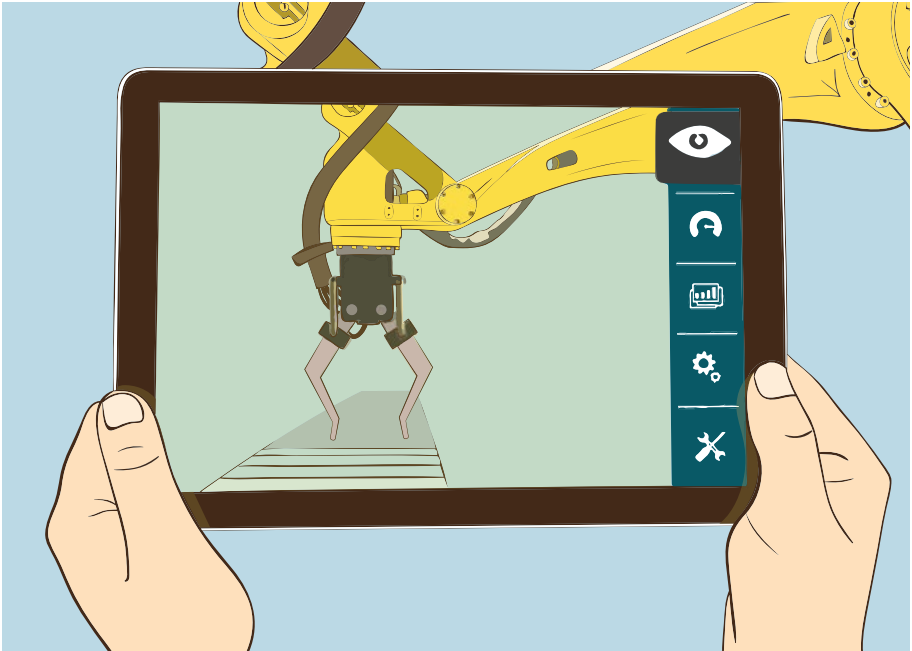
Unidos), Mooverang y Fintonic utilizan nuestro histórico de datos para proponernos métodos de ahorro, retos, monitorización de nuestras cuentas o servicios de alertas automáticos. Estas empresas no generan los datos, sino que nos piden permiso a los usuarios para acceder a ellos y a partir de ahí derivar toda la información. Todo un ejemplo de lo que es la Economía de los Datos.

2.5

Obstáculos y adopción del *big data* en España y América Latina

Los valores del *big data* y su impacto en la economía y los resultados de las empresas es directo y, en la actualidad, la mayoría son conscientes de su valor. Sin embargo, en las mismas existen obstáculos que hacen que algunas no estén abordándolos con toda la potencia posible:

- Entendimiento de qué puede resolver: como indicábamos anteriormente, el reto principal es traducir los problemas de negocio en problemas de datos, y eso es algo que, hoy en día, cuesta realizar internamente en las compañías.
- No es sencillo: los proyectos con datos no son inmediatos, tienen un componente científico y exploratorio que hace que la solución no siempre sea una línea recta. Debido a esta incertidumbre, es posible que determinados casos que se planteen no sean viables o que los proyectos, así como sus costes asociados, se alarguen en el tiempo.
- No se encuentra el talento necesario: por la mezcla de aptitudes requeridas para realizar proyectos de *data science*, existe poco personal cualificado y, normalmente, las empresas no son capaces de generar estas capacidades.
- Es complicado estimar el retorno: en muchos casos de uso, es difícil estimar el retorno de la inversión en un proyecto.
- Aspectos regulatorios: determinados casos de uso o políticas de datos internas en una empresa pueden tener consecuencias legales que hay que tener en cuenta previamente, lo que es un aspecto fundamental para algunas de las iniciativas relacionadas con datos.
- Del análisis al trabajo del día a día: las empresas que ya están haciendo proyectos de *big data* consideran que es complicado llegar a operativizar los resultados que de ellos obtienen. En muchas ocasiones, los resultados de estos proyectos son conclusiones de negocio que se pueden convertir en acciones estratégicas pero, en otros casos, el resultado debe integrarse en los sistemas de información de las empresas, lo que requiere su mantenimiento, adaptación y evolución, tareas que aumentan el coste de los proyectos y, en ocasiones, los frenan.



Por todo ello, siendo una de las tendencias estratégicas que con más fuerza está irrumpiendo en empresas de todo tipo de sectores, el proceso de adopción y madurez está todavía en curso.

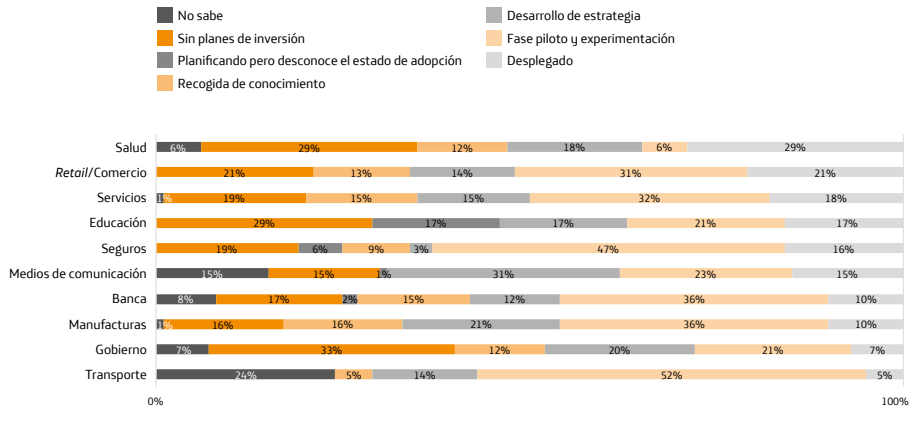
Según distintos informes, en España, el 84,8% de las compañías está realizando proyectos *big data* o tiene planes inminentes para hacerlo.³ En todo el mundo, si analizamos cómo está siendo esta adopción por sectores, el 70% de las empresas financieras declara que ya han hecho este tipo de proyectos, enfocados, principalmente, en el conocimiento profundo de los clientes, la mejora de sus modelos de propensión a la adopción de sus productos, el análisis de riesgos o la detección de fraude de diferentes tipos, como en transacciones con tarjetas de pago o peticiones de préstamos. En el caso del sector telecomunicaciones, aproximadamente el 60% de las organizaciones está realizando proyectos *big data* para analizar la movilidad de sus usuarios, entender cómo están conectados, mejorar sus campañas de *marketing* o realizar comunicaciones personalizadas gracias a la información geográfica que tienen. Los medios de comunicación, un sector donde la transformación digital ha hecho que más se hayan modificado los modelos de negocio tradicionales por la aparición de los portales de noticias online, el principal uso del *big data* está relacionado con la recomendación personalizada de contenidos y la creación de nuevos productos basados en datos, y un 73% de estas empresas declara estar inmerso en este tipo de proyectos. En otros sectores, como *retail* o logística, la adopción del *big data* supera ampliamente el 50%.⁴

3. «Situación de las TIC en España», <https://www.esmartcity.es/2013/01/15/situacion-de-las-tic-en-espana>

4. «Industries using big data», <https://upxacademy.com/big-data-use-cases-industries/> (Poddar 2016) «State of big data adoption», <https://www.slideshare.net/denisreimer/big-data-industry-insights-2015>

Fig. 2.6

Estado de la adopción del *big data* por industria



Fuente: Gartner, «State of big data adoption», 2015, disponible en www.slideshare.net/denisreimer/big-data-industry-insights-2015.
 Nota: apenas el 14% de la muestra afirma haber realizado el despliegue de capacidades *big data*. El 44% en estado de pilotaje o desplegando capacidades.

Por otro lado, en América Latina, el fenómeno del *big data* también es imparable y las inversiones en este tipo de proyectos crece año tras año. Los países donde más se está apostando por el desarrollo de esta tecnología son Brasil, México y Argentina, donde más de un 75% de las empresas declara que es estratégico para ellas, está realizando proyectos o tiene planes para llevarlos a cabo en los dos próximos años, con un mercado estimado total superior a los seis mil quinientos millones de dólares.⁵

En definitiva, el *big data* tiene un componente transversal en todos los sectores económicos y en empresas de todo tipo y tamaños y, desde todos los ámbitos, las organizaciones están ya extrayendo valor de sus datos. Esto implica que, hoy en día, existe una gran necesidad de profesionales cualificados en *big data* y *data science* y que, al ser un campo relativamente reciente, no se tengan las capacidades suficientes, en muchos casos, para poder desarrollar todas las posibles aplicaciones que impactarían en el negocio de las empresas.

5. (Bowden 2014) «The Key Findings on big data Study», <http://www.digital-warriors.com/key-findings-big-data-study/>
 (Carreño 2014) «El *big data* crecerá un 657% en LATAM», <http://pulsosocial.com/2014/05/30/big-data-crecera-un-657-en-los-proximos-5-anos-en-latinoamerica/>

3

La regulación de los datos en el mundo

3.1

La protección de los datos personales en el mundo

54

3.2

El enfoque europeo

55

3.3

América Latina

60

La esencia de la regulación de los datos es la salvaguarda de la privacidad de información sensible, generalmente mediante la prohibición de la divulgación o el uso indebido de información sobre personas físicas. La preocupación por la protección de los datos comenzó en los años sesenta en Estados Unidos y Europa (Suecia, Alemania) en paralelo a los avances informáticos, el procesamiento automático y el almacenamiento en soportes magnéticos de datos personales.

La recopilación y el almacenamiento generalizados de información personal por parte de las autoridades —censo, identificación, huellas, fotografías, declaraciones de impuestos—, servicios de salud —datos médicos e historiales clínicos— o de organizaciones privadas como entidades financieras —gastos personales, transacciones de pago, centrales de riesgo— contribuyó a la necesidad de dotar de reglas de juego al tratamiento de datos personales. El desarrollo de las TIC, especialmente Internet, y las posibilidades de transmitir información sensible de manera masiva han crecido exponencialmente.

Cuando las primeras regulaciones fueron diseñadas, no tuvieron en cuenta la entonces impredecible irrupción de Internet en todos los aspectos de la vida de la gente, ni que cada persona estaría equipada con un dispositivo conectado en forma de teléfono, con capacidad de capturar y enviar datos permanentemente. El inminente desarrollo del Internet de las Cosas (IoT), sustentado en la captura y el envío autónomos de datos por parte de objetos cotidianos conectados a Internet, además de implicaciones relativas a la seguridad, también desvela nuevas preocupaciones relativas a la privacidad y la protección de datos, en este caso no necesariamente de carácter personal. No en vano, los datos han sido descritos como «la contaminación de la era de la información, y la protección de la privacidad, su reto medioambiental».¹

Europa es la región del mundo en que la protección de datos ha alcanzado un nivel más elevado, mientras que en Norteamérica es la privacidad la que ha alcanzado un alto grado de desarrollo. En otras regiones se han producido importantes avances en materia de legislación e institucionalización de la protección de la privacidad y la protección de los datos personales. La institucionalización responde a la necesidad de que las leyes que regulan el tratamiento de los datos personales estén acompañadas de autoridades de supervisión encargadas del control de su fiel cumplimiento. Así, el análisis de la idoneidad y la adecuación de la protección debe comprender estos dos elementos básicos: el contenido de las normas aplicables y los medios para asegurar su aplicación eficaz.

1. (Buttarelli 2015) https://edps.europa.eu/sites/edp/files/publication/15-12-10_edemocracy_en.pdf

3.1

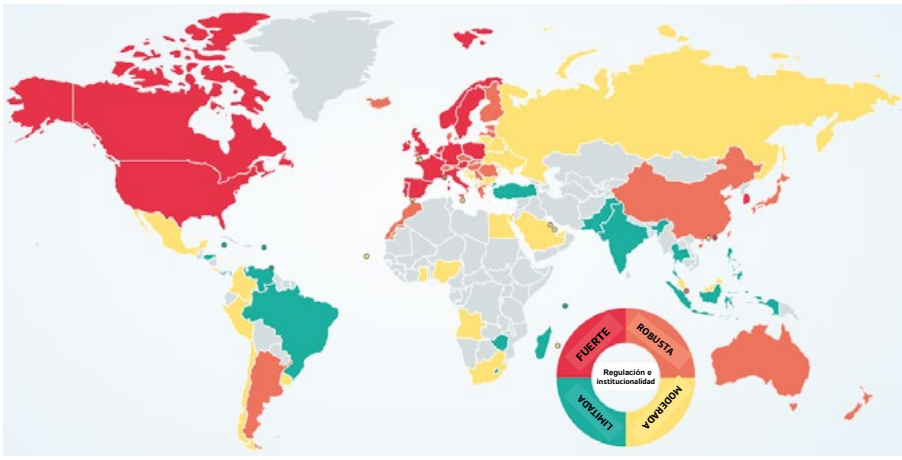
La protección de los datos personales en el mundo

Internet y la digitalización de bienes y servicios han transformado la economía mundial, de manera que la transferencia de datos, incluidos los de carácter personal, figura entre las actividades cotidianas de las empresas. Puesto que los intercambios comerciales dependen cada vez más de la circulación de datos personales, su protección y su seguridad constituyen un elemento fundamental de la confianza de los consumidores.

Un creciente número de países ha adoptado o está elaborando nuevas leyes en materia de protección de datos y privacidad. Solo en 2015, 109 países contaban con leyes promulgadas (76 en 2011), y en 2017 son 35 los que están en proceso de elaborar legislación. En general, estas «nuevas» legislaciones contemplan en mayor o menor medida un conjunto básico de principios comunes en la regulación de los datos, tales como: (i) el reconocimiento de la protección de datos como derecho fundamental, (ii) la adopción de legislación general, (iii) la existencia de derechos individuales y exigibles en materia de privacidad y (iv) la creación de una autoridad de control independiente.

Fig. 3.1

La regulación de los datos en el mundo



Fuente: 2017 DLA Piper.

3.2

El enfoque europeo

En el ámbito europeo, la protección de datos está configurada como un derecho fundamental de los ciudadanos. La principal norma en la materia es la Directiva 95/46 de la Unión Europea, aunque ya en 1981 el Consejo de Europa adoptó el Convenio n.º 108, sobre la protección de las personas, el único instrumento internacional vinculante sobre protección de datos.

El Estándar Internacional para la Protección de la Privacidad y la Información Personal (Resolución de Madrid) fue adoptado el 9 de mayo de 2009 y entró en vigor el 1 de junio de 2009. Respondía a la urgencia expresada por las autoridades de protección de datos de Suiza y España, respaldada por otras veinte autoridades² en la 30ª Conferencia Internacional de Autoridades Protección de Datos y Privacidad, sobre la necesidad de proteger la privacidad en un mundo sin fronteras, y de alcanzar una propuesta conjunta de estándares internacionales. El estándar consta de seis principios básicos —lealtad y legalidad; finalidad; proporcionalidad; calidad; transparencia, y responsabilidad— y cuatro derechos básicos del interesado, los denominados derechos ARCO —acceso, rectificación, cancelación y oposición—. Los estándares buscan adecuar la regulación a la realidad motivada por los incesantes desarrollos tecnológicos y la globalización de los intercambios de datos y, en el caso europeo, para garantizar un adecuado despliegue del Mercado Único Digital.

Los principios y derechos básicos contemplados en el estándar internacional quedan recogidos en el Reglamento General de Protección de Datos (RGPD) de la Unión Europea, aplicable a partir de mayo de 2018 como normativa única paneuropea, cuya elaboración arrancó en 2012. Con la creación de un mecanismo de ventanilla única, corresponderá a una sola autoridad de protección de datos controlar las operaciones transfronterizas de tratamiento de datos que las empresas lleven a cabo en la Unión Europea. En este contexto, las empresas domiciliadas fuera de la Unión Europea tendrán que aplicar las mismas normas que las europeas si suministran bienes y servicios o vigilan la conducta de los ciudadanos en la Unión Europea.

El RGPD extiende los derechos de los interesados actualmente vigentes (derechos ARCO) a cuatro nuevos derechos (portabilidad, oposición a decisiones automatizadas, limitación y supresión). En esta línea, tras la sentencia del 13 de mayo de 2014 del Tri-

2. Comisario Federal de Protección de Datos (Alemania), Comisario de Protección de Datos y Libertad de Información de Berlín (Alemania), Comisario de Protección de Datos de Schleswig-Holstein (Alemania), Comisión de Protección de Datos (Austria), Comisión de Protección de la Privacidad (Bélgica), Comisión de la Informática y las Libertades (Burkina Faso), Comisario de Privacidad (Canadá), Comisario de Acceso a la Información (Canadá), Supervisor Europeo de Protección de Datos, Comisario de Información (Eslovenia), Agencia Española de Protección de Datos (España), Agencia Catalana de Protección de Datos (España), Agencia de Protección de Datos de la Comunidad de Madrid (España), Agencia Vasca de Protección de Datos (España), Comisión Nacional de la Informática y las Libertades (Francia), Comisario de Privacidad para la Protección de Datos (Hong Kong), Comisario de Protección de Datos (Irlanda), Garante para la Protección de los Datos Personales (Italia), Comisario de Privacidad (Nueva Zelanda), Comisión de Protección de Datos (Países Bajos), Comisión Nacional de Protección de Datos (Portugal), Comisario de Información (Reino Unido), Oficina para la Protección de Datos Personales (República Checa), Comisario Federal de Protección de Datos (Suiza).

bunal de Justicia Europeo, se aplica en Europa el derecho al olvido, por el cual se reconoce el derecho a impedir la difusión de información personal en buscadores generales cuando la información es obsoleta o ya no tiene ni relevancia ni interés para el público.

Fig. 3.2

Derechos del Reglamento General de Protección de Datos (RGPD)



Fuente: Afi.

El RGPD establece importantes sanciones económicas por incumplimiento de la norma a empresas, Administraciones públicas y organizaciones, pudiendo llegar a los veinte millones de euros o el 4% de la facturación global anual.

Fig. 3.3

Sanciones del Reglamento General de Protección de Datos (RGPD)



Tras la aprobación del RGPD, la Comisión Europea ha propuesto una nueva legislación en materia de privacidad de las comunicaciones electrónicas (Directiva de Privacidad y Comunicaciones Electrónicas o ePrivacidad)³ tras su revisión a inicios de 2017, con el objetivo de convertirla en reglamento para crear junto con el RGPD un nuevo marco unitario en toda la Unión Europea. La revisión adapta las normas aplicables a las comunicaciones online al nuevo RGPD, a las que estarán sujetos los operadores de telecomunicaciones tradicionales y los nuevos proveedores de servicios de comunicaciones electrónicas de transmisión libre u *over the top* no contemplados en la anterior versión, esto es, servicios y contenidos, principalmente de audio y vídeo, que se transmiten a través de infraestructuras de banda ancha, sin que las operadoras y sus redes puedan controlar su distribución (servicios de música y vídeo, de mensajería instantánea, e incluso de llamadas de voz sobre IP (VoIP) entran en esta categoría.

La nueva normativa de ePrivacidad agudiza las medidas de seguridad exigibles de los equipos terminales (incluye las transmisiones de máquina a máquina o IoT), el contenido de las comunicaciones y la eliminación o anonimización de metadatos, por señalar algunos ejemplos. En relación con las comunicaciones comerciales, se explicita la obligación de obtención del consentimiento previo y expreso para su envío a usuarios finales (en la línea de lo establecido en el RGPD), así como otras obligaciones como la identificación del tipo de llamada (por ejemplo, prefijo identificativo de llamada comercial). Asimismo, contempla que los Estados miembros podrán conceder a los usuarios la opción de incluirse en listas que declaran el deseo de no recibir ningún tipo de comunicación comercial. En relación con la información y datos obtenidos a través de *cookies*, se establece que no será necesaria la aceptación del usuario en todas las páginas web a las que accede, sino que los usuarios podrán realizar una configuración general de la privacidad, exceptuando las *cookies* consideradas no invasivas, como las que mejoran la navegación o realizan medición de visitas.

En este nuevo contexto regulatorio, el conjunto de industrias se encuentra en proceso de adaptación al nuevo marco normativo, si bien queda pendiente en muchos casos desarrollar metodologías estandarizadas para dar respuesta a las solicitudes de ejercicio de derechos, especialmente cuando se ofrecen bienes y/o servicios por medios electrónicos. Por su parte, los proyectos de *big data* deben tener en cuenta el marco jurídico aplicable (RGPD y ePrivacidad) a partir del 25 de mayo de 2018, aunque no sea de aplicación en los siguientes casos:

- cuando no sea posible la identificación de los individuos (anonimización),
- cuando la identificación requiera esfuerzos desproporcionados y
- cuando los datos tratados y analizados no sean datos personales (datos relativos al tráfico, la contaminación o el clima).

3. Propuesta de Reglamento del Parlamento Europeo y del Consejo sobre el respeto de la vida privada y la protección de los datos personales en el sector de las comunicaciones electrónicas y por el que se deroga la Directiva 2002/58/CE.

Sí se aplica a todos los datos personales con independencia de que estos sean primarios o metadatos (visibles o invisibles).

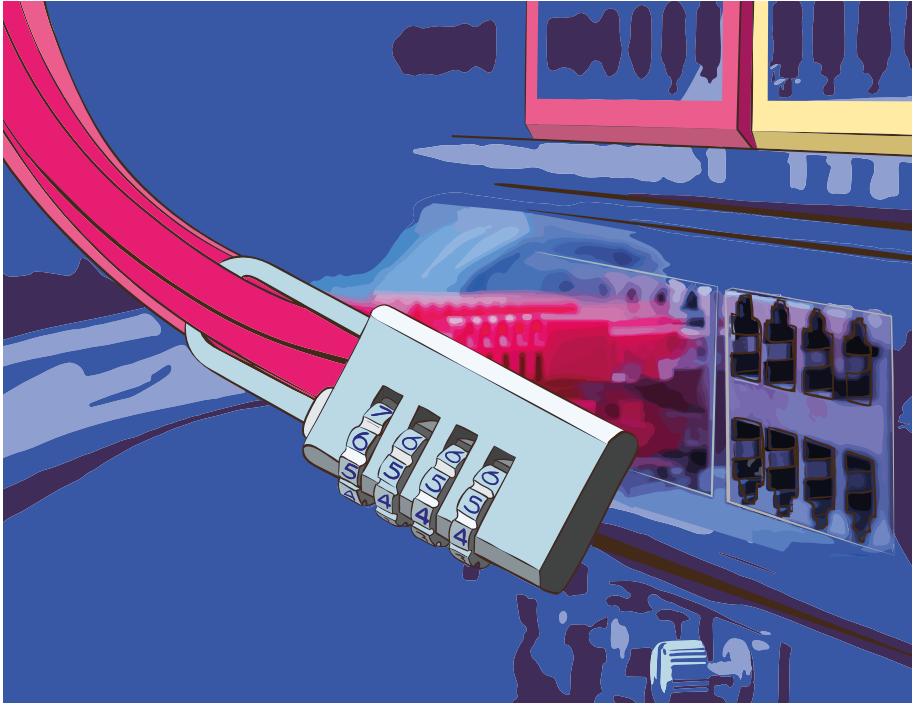
La Directiva 95/46 —vigente hasta la entrada en vigor del RGPD— prohíbe la transferencia de datos a países que no cuenten con un nivel adecuado de protección de datos (principio de adecuación), y establece un procedimiento para determinar formalmente si un país ofrece dicho nivel de protección, contenido en un dictamen de 1998.

La principal consecuencia de que un país sea declarado adecuado es que se podrán transferir datos desde los Estados miembros de la Unión Europea sin necesidad de ningún tipo de trámite o autorización especial. Hasta la fecha, han sido considerados tales Suiza, Andorra, las islas Feroe, Guernsey, Jersey y la Isla de Man (países estrechamente vinculados con la Unión Europea y sus Estados miembros), Argentina, Canadá, Israel y Estados Unidos (importantes socios comerciales) y Nueva Zelanda y Uruguay, pioneros en la elaboración de leyes de protección de datos en sus respectivas regiones.

En los casos de Canadá y Estados Unidos, se trata de adecuaciones parciales. En Canadá solo es aplicable a entidades privadas comprendidas en la Personal Information Protection and Electronic Documents Act y, en Estados Unidos, el Escudo de Privacidad Unión Europea-Estados Unidos (*Privacy Shield*) es el instrumento habilitado ante la falta de una legislación general estadounidense en materia de protección de datos, por medio del cual las empresas participantes asumen los compromisos establecidos en dicho régimen. Cabe recordar que el Tribunal de Justicia de la Unión Europea (TJUE) declaró inválida la Decisión de la Comisión Europea 2000/520/CE que establecía el nivel adecuado de protección de las garantías para las transferencias internacionales de datos a Estados Unidos ofrecidas por el acuerdo de Puerto Seguro. En julio de 2016, la Comisión Europea publicó la Decisión 2016/1250 sobre la adecuación de la protección conferida por el nuevo *Privacy Shield*.

En relación con Reino Unido, una vez se materialice su salida de la Unión Europea, los planes desvelados por el Gobierno británico sugieren la aprobación de una nueva ley de protección de datos que sustituya la British Data Protection Act 1998, y que recoja muchos de los puntos fundamentales del RGPD para asegurar que el flujo de datos entre la Unión Europea y Reino Unido pueda continuar una vez consumado el Brexit.

En enero de 2017, la Comisión Europea emitió una comunicación al Parlamento Europeo y al Consejo de Europa sobre «intercambio y protección de los datos personales en un mundo globalizado», señalando que «la solidez de la protección de datos facilita su circulación, al fomentar la confianza de los consumidores en aquellas empresas que se preocupan por el tratamiento que dan a los datos personales de sus clientes». De esta forma, la Comisión Europea defiende, por un lado, que el cumplimiento de unas estrictas normas de protección de datos supone una ventaja en la Economía Digital mundial. Por otro, que la cooperación en materia de aplicación de la ley redundará en un rápido y eficaz intercambio de información para luchar contra la delincuencia en sus formas tradicionales y nuevas.



Con esta motivación, la Comisión Europea solicita a la Unión Europea que colabore activamente con terceros países para lograr una convergencia al alza de los principios de protección de datos a escala internacional. En algunos casos, anticipa la Comisión Europea, podrá resultar más conveniente recurrir a una adecuación parcial o sectorial (por ejemplo, en servicios financieros o tecnologías de la información), de zonas geográficas o sectores industriales que representan una parte importante de la economía de un determinado tercer país.

No son muchos los países fuera de Europa que cuentan con autoridad de protección de datos reconocida por la Conferencia Internacional de Comisionados de Protección de Datos y Privacidad. En particular, son cinco en África (Burkina Faso, Marruecos, Mauricio, Senegal y Túnez), ocho en el continente americano (Estados Unidos, Canadá, México, Costa Rica, Colombia, Perú, Uruguay y Argentina), tres en Asia (Hong Kong, Singapur y Corea del Sur), además de Australia y Nueva Zelanda.

En España, la Agencia Española de Protección de Datos (AEPD), creada en 1992 y operativa desde 1994, y las agencias autonómicas de Cataluña (2003) y País Vasco (2004), ejercen las funciones de control respecto de los ficheros de datos de carácter personal creados o gestionados por las comunidades autónomas y por la Administración local de su ámbito territorial. Los ficheros privados de estas comunidades autónomas, no obstante, son competencia de la AEPD. En 2017, la AEPD publicó el Código de Buenas Prácticas en Protección de Datos para Proyectos *Big Data*.

3.3

América Latina

La Red Iberoamericana de Protección de Datos (RIPD) es resultado del acuerdo alcanzado en el Encuentro Iberoamericano de Protección de Datos en La Antigua (Guatemala) en 2003, foro de promoción del derecho a la protección de datos en la comunidad iberoamericana. En junio de 2017, en el marco del XV Encuentro Iberoamericano de Protección de Datos, la RIPD aprobó los llamados «Estándares de Protección de Datos de los Estados Iberoamericanos».⁴

A pesar de que los países latinoamericanos han decidido inspirarse en el RGPD, América Latina se encuentra por detrás de los estándares europeos y estadounidenses; ello se debe a que las leyes actuales fueron elaboradas con la referencia de la Directiva Europea de Protección de Datos de 1995, que no contempla muchos aspectos hoy considerados básicos en la economía y en la sociedad digitales. Algunos países, como es el caso de Brasil, no cuentan con una ley de protección de datos completa que dote de confianza a los agentes, y otros, así sucede en Argentina o Chile, cuentan con leyes obsoletas. Dicho esto, los tres están elaborando nuevos marcos normativos para conseguir su adecuación.

Colombia, Perú y Uruguay se encuentran más avanzados que los anteriores en materia de seguridad de los datos y notificación, aunque con importantes vacíos en relación con los mecanismos de reporte de incidentes a las autoridades. En materia de ciberseguridad y de definición de una estrategia nacional, solo en Chile, Colombia o Paraguay se han producido algunos avances.

Es urgente que los países de América Latina actualicen debidamente su legislación en materia de protección de datos personales y establezcan mecanismos de ciberseguridad para proteger a sus ciudadanos en el nuevo contexto digital en que se encuentra este mundo globalizado.

Tabla 3.1 Normativa y autoridades de protección, supervisión de datos y privacidad en América Latina

	Normativa	Autoridad
Argentina	Ley N° 25.326 de protección de datos personales de 2000, modificada por Ley 26.343 de 2008 y reglamentada por Decreto N° 1558/01, modificado por Decreto N° 1160/10.	Dirección Nacional de Protección de Datos Personales (DNPDP).

4. Tomando como referencia otros instrumentos internacionales emblemáticos en materia de protección de datos personales, como las directrices relativas a la protección de la intimidad y de la circulación transfronteriza de datos personales de la OCDE; el Convenio número 108 del Consejo de Europa para la protección de las personas con respecto al tratamiento automatizado de datos de carácter personal y su protocolo; el Marco de Privacidad del Foro de Cooperación Económica Asia Pacífico, y el RGPD, entre otros.

Chile	<p>Ley N° 19.628 Sobre protección de la vida privada y Protección de Datos de Carácter Personal (1999).</p> <p>Reglamento del Registro de Bancos de Datos Personales a cargo de Organismos Públicos.</p> <p>Proyecto de Ley N° 395-359 en Congreso Nacional.*</p>	Servicio de Registro Civil e identificación / Modernización y Gobierno Digital.
Colombia	Ley 1581 de Protección de Datos Personales (2012).	Defensoría del Pueblo / Superintendencia de Industria y Comercio.
Costa Rica	<p>Ley 8968 de Protección de la Persona frente al tratamiento de sus datos personales (noviembre 2011, en vigor 2016).</p> <p>Decreto Ejecutivo 37554, Reglamento de marzo 2013, reformado mediante Decreto Ejecutivo 40008 - Derecho a la Autodeterminación Informativa (ARCO).</p>	Agencia de Protección de Datos de los Habitantes.
El Salvador	Decreto 695 Ley de regulación de los servicios de información sobre el historial de crédito de las personas (julio de 2011).	Superintendencia General de Electricidad y Telecomunicaciones.
México	<p>Ley Federal de Protección de Datos Personales en Posesión de los Particulares y se reforman los artículos 3, fracciones II y VII, y 33, así como la denominación del Capítulo II, del Título Segundo, de la Ley Federal de Transparencia y Acceso a la Información Pública Gubernamental (julio 2010).</p> <p>Ley General de Protección de Datos Personales en Posesión de Sujetos Obligados (enero de 2017).</p>	<p>Instituto Federal de Acceso a la Información y Protección de Datos.</p> <p>Instituto de Acceso a la Información Pública y Protección de Datos Personales del Distrito Federal.</p>
Nicaragua	Ley n.º 797 de Protección de Datos Personales (marzo de 2012).	Dirección de Protección de Datos Personales adscrita al Ministerio de Hacienda y Crédito Público.
Perú	<p>Ley n.º 29733 de Protección de Datos Personales (julio de 2011).</p> <p>Decreto Legislativo N° 1353 que crea la Autoridad Nacional de Transparencia y Acceso a la Información Pública y modifica la Ley N° 29733, Ley de Protección de Datos Personales (enero de 2017).</p>	Dirección General de Transparencia, Acceso a la Información Pública y Protección de Datos Personales, órgano que depende jerárquicamente del Despacho Viceministerial de Derechos Humanos y Acceso a la Justicia.
Uruguay	<p>Ley N° 18.331 de Protección de Datos Personales y Acción de Habeas Data (LPDP) (agosto 2008).</p> <p>Reglamento Decreto n.º 414/009 (agosto de 2009).</p>	Unidad Reguladora y de Control de Datos Personales (URCDP).

Fuente: Organización de Estados Americanos (OEA).

* En 2010, Chile firmó el Convenio de Adhesión a la OCDE y ello supuso la asunción de la Recomendación sobre Protección de la Privacidad y Flujo Transfronterizo de Datos Personales, materia ausente en la Ley N° 19.628. Por dicho motivo, ingresó en noviembre de 2011 al Congreso Nacional el Proyecto de Ley N° 395-359 que modifica la actual ley.

Parte II

Mercados y oportunidades

4

El mercado de los datos

4.1

Agentes

67

4.2

Ciclo de vida del dato: generación, captura, transmisión, almacenamiento, análisis, procesamiento

72

4.3

¿Dónde se almacena el *big data*?

80

4.4

El análisis de datos y sus profesionales: estado actual, demanda actual, *gap* y previsiones futuras

81

4.5

Evolución del volumen de datos y su utilización en la Economía Digital

85

4.6

***Open data*: concepto, limitaciones y medidas adoptadas en favor de la creación de estándares**

88

4.1

Agentes

La Economía de los Datos no es solamente un hecho que ha surgido en las empresas, sino toda una estrategia encaminada a extraer valor económico de la información que ellas mismas generan. Los nuevos grandes gigantes tecnológicos como Google o Facebook basan toda su generación de valor en la explotación de los datos que ellos mismos crean a través de un servicio, ya sea la búsqueda de información en Internet o la conexión con nuestros amigos.

Cada vez que realizamos una búsqueda, escribimos un email, comentamos una foto o clicamos un *like* en alguna marca, estas empresas están almacenando toda nuestra traza, generando una completa historia digital de nuestra vida. Se convierten, a través de un servicio que ofrecen a cientos (y ya llegan a miles) de millones de usuarios, en generadores de datos que, posteriormente, convertirán en ganancias económicas. ¿Cómo lo hacen?

Uno podría pensar que simplemente comparten sus bases de datos con otras empresas, integrándolas en sus sistemas de información, pero esta práctica presenta multitud de trabas tecnológicas y complejidades legales que no permiten hacerlo. En realidad, lo que estas empresas hacen es crear otro *software*, otro servicio que, apoyándose en nuestros datos, resuelven un problema adicional a otro tipo de usuarios y/o clientes. Si el cliente original de un servicio online es el usuario (pensemos en Google o Facebook de nuevo, cuyos usuarios son, potencialmente, cualquier persona con conexión a Internet), el cliente del nuevo servicio es, en general, una empresa que utiliza el *software* desarrollado, alejando a este nuevo cliente del dato bruto real para diferentes propósitos: *marketing* más personalizado, análisis de la movilidad, posicionamiento de marca y publicidad, estudio de la inversión en medios digitales, etc., pero nunca lo hace accediendo al dato original que la compañía recogió.

Este ejemplo es suficiente para identificar que, en función de su posición dentro de la cadena de valor del dato, pueden encontrarse diferentes actores en el mercado de los datos:

- Generadores de datos: cualquier empresa que, ya sea como fin último o no, genera datos a través de su operativa, susceptibles estos de ser aprovechados por otros. En una definición tan genérica caben tanto Gobiernos y Administraciones públicas (que generan datos a través de encuestas, como en el caso de los institutos de estadística, a través de sensores, o simplemente a través de las partidas presupuestarias que asignan a los diferentes departamentos gubernamentales), como universidades y organizaciones científicas (por ejemplo, en el ámbito de la investigación física, se genera una cantidad ingente de datos procedentes de los experimentos o, en investigación médica, toda la información de los pacientes sobre los que se investiga) y empresas.

- Mercados de datos: plataformas donde distintos agregadores han desarrollado sistemas de integración de diferentes fuentes de datos públicas y no públicas. Como se verá posteriormente, uno de los principales retos que tienen las empresas es, precisamente, la transformación de los datos, internos y externos, para tratarlos de manera unificada y estándar. Estas empresas se encargan precisamente de realizar esta unificación de manera que se puedan consultar estas fuentes como si procedieran de una única. Su modelo de negocio —cómo generan beneficios— suele tener una componente *freemium*, es decir, algunos de estos datos pueden ser consultados de manera gratuita, pero para la gran mayoría, especialmente para los datos que mayor valor tienen, es necesario que las empresas paguen por ello. Empresas como Knoema y datamarket.com (adquirido recientemente por la compañía desarrolladora de un *software* de análisis de datos llamado QlikView) son ejemplos de estos mercados de datos.
- Empresas tecnológicas: empresas que se dedican tanto a crear sus propias tecnologías de almacenamiento y procesamiento de la información como a implantar en compañías otras tecnologías *open source* (de código abierto, *software* distribuido y desarrollado libremente) disponibles en el mercado. También suelen tener entre sus líneas de servicio la habilitación de las infraestructuras adecuadas para mantener la operativa de todos estos servicios.
- Compañías usuarias de datos: aunque la mayoría de empresas hoy en día están muy enfocadas en extraer valor de sus datos internos, cada vez más empiezan a mirar hacia fuera para identificar qué disponibles pueden complementar los que ya tienen para, de esta manera, diseñar procesos de análisis y modelos predictivos mucho más ricos. En este sentido, prácticamente cualquier compañía es una potencial usuaria de servicios de terceros.
- Compañías de servicios analíticos: puesto que el *big data* y la *data science* son un conjunto de técnicas y tecnologías que comprenden una mezcla de aptitudes entre las ciencias de la computación, la estadística o las matemáticas y, dado que su explosión se ha producido en los últimos años, muchas veces las empresas no cuentan con los recursos suficientes para llevar a cabo estos proyectos. Es por ello que empresas de servicios profesionales se han especializado en la implantación de plataformas analíticas, la formación y la integración de equipos especializados en aprendizaje automático —la rama del modelado predictivo dentro de la *data science*— o la visualización, una rama cuyo objetivo es comunicar los resultados de los análisis de una manera atractiva.
- Compañías de *software* basado en datos: finalmente, otras compañías han optado por construir productos de *software* sobre datos, ya sea de datos de terceras empresas o producidos por otro tipo de organismos. En este sentido, podemos encontrar negocios verticalizados en sectores concretos como finanzas, salud, ámbito legal o *marketing*.

Recuadro 4.1 LUCA: *Data-driven decisions*

LUCA (nombre prestado del antepasado universal común más reciente de todo el conjunto de los seres vivos actuales, *last universal common ancestor*) es la nueva unidad de servicios *big data* de Telefónica para empresas y todo tipo de organizaciones públicas y privadas de más de veinte países, lanzada tras la adquisición a finales de 2015 de la empresa Synergic Partners. LUCA cuenta con una red de más de cien especialistas en *data engineering*, infraestructura, *data science* y *business insights*, cuya propuesta de valor transita en estos tres ámbitos.

- Extracción de valor de los datos anonimizados y agregados de las redes de Telefónica, incluyendo servicios ya existentes como Smart Steps, enfocado en soluciones de gestión de la movilidad para una planificación más eficiente de, por ejemplo, los servicios de transporte y los servicios turísticos en las ciudades o, en el caso de emergencias sanitarias, con capacidad de ayudar a entender mejor los movimientos de la población y limitar la propagación de pandemias.
- Servicios de analítica y consultoría externa a clientes nacionales e internacionales que ofrece Synergic Partners.
- Optimización de la gestión y extracción de valor de los propios datos de la organización utilizando las infraestructuras en la nube de Telefónica, en un esquema BDaaS (*big data as a service*).

El catálogo de productos y servicios de LUCA es relevante para una amplia gama de sectores, tales como *retail*, turismo, transporte, sector público, servicios financieros y publicidad exterior, como así reflejan las siguientes experiencias:

En el ámbito del turismo, LUCA ha desarrollado proyectos en estos países:

- Brasil: extracción de *insights* del sector turístico del Estado de Espírito Santo.
- España: análisis *data-driven* de visitantes del Museo Reina Sofía y del Palacio Real de Madrid; mejora de la experiencia de los esquiadores en Sierra Nevada; comprensión del turismo en la ciudad de Valencia durante Las Fallas o en la ciudad de Gerona durante el festival anual Temps de Flors, e impacto turístico de las carreras de Fórmula 1 y Moto GP del circuito de Montmeló en la zona.
- México: obtención de métricas clave de turistas del Municipio de León durante el Festival del Globo para optimizar los procesos de decisión municipal de cara a otras ediciones.
- Perú: posicionamiento de la imagen de Promperú.

En transporte (LUCA Transit), destacan estas otras:

- En Reino Unido, la optimización del plan de carreteras de Highways England, empresa responsable de operar, mantener y mejorar la red de carreteras de Reino Unido, así como el estudio de cuota de mercado para la compañía ferroviaria East Coast Main.

- En Estados Unidos, con los ayuntamientos de Newark y Sherwood se proporcionó información detallada sobre los patrones de desplazamiento de los vehículos y el perfil de los conductores.
- En Ecuador, la colaboración entre el municipio de Quito y el Banco Interamericano de Desarrollo (BID), para conocer los comportamientos y movimientos de los usuarios del metro de Quito.
- En Colombia, la colaboración con el Ministerio de Asuntos Exteriores de Reino Unido, Atkins, FDI Pacífico y la ciudad de Cali, para evaluar e identificar oportunidades en el proyecto de renovación urbana Corredor Verde.
- En la municipalidad de Neuquén —área de concentración más grande de la Patagonia argentina—, estudio de los flujos de la movilidad integral de la ciudad, utilizado para el diseño del nuevo sistema de transporte público de pasajeros y la traza del nuevo metrobús inteligente.
- En Brasil, en colaboración con el Banco Mundial y la Universidad de São Paulo, recogida de datos de los abonos de transporte, datos móviles de app y datos de una lista de proveedores de telecomunicaciones, para investigar los patrones de movilidad de residentes en Paraisópolis.
- En España, en colaboración con el Consorcio de Transportes del Área de Zaragoza, uso de datos de LUCA, para crear matrices para determinar la demanda de transporte en la ciudad y tendencias demográficas.

En el ámbito del *retail* (LUCA Store):

- En Ecuador, desarrollo de solución integral que cuenta con tres herramientas de recolección de información de los clientes: LUCA Store para obtener información de flujos de movilidad dentro de las ciudades; contadores 3D para identificar las personas que entran a los centros comerciales, y wifi inteligente para ubicar a las personas dentro del centro comercial e identificar zonas calientes.
- En Reino Unido, junto con Morrisons —cadena de supermercados—, uso de tecnología *small cell*, para comprender la digitalización de los clientes, la afluencia, las visitas periódicas y los análisis de métodos de captación. *Small cell* también se ha utilizado para ayudar a O2 UK a optimizar su huella comercial.
- En Argentina, colaboración con Netshoes, para promocionar su app móvil durante el periodo de *Black Friday*; se dió acceso a los clientes de Vivo a su app y sitio web sin consumo de datos.
- En Brasil, decisión *data-driven* para la ubicación de establecimientos de Natura en la ciudad de São Paulo, con la plataforma Smart Steps.

En publicidad exterior, LUCA Audience ha trabajado con Exterior Media facilitando la transformación digital en su estrategia de datos. En servicios financieros, LUCA ha apoyado a Bradesco, banco líder en Brasil, a migrar sus clientes de los canales tradicionales a la aplicación móvil del banco con el producto Sponsored Data, consiguiendo así posicionar a la app de Bradesco como la más descargada en Brasil. También ha contribuido a reducir el fraude *Account Takeover* para Lloyds Bank.

Fuente: Afi, a partir de datos facilitados por Telefónica.

Además, claves para la consolidación del *big data* como un campo maduro en el ámbito empresarial y público, cabe destacar los siguientes actores en el mercado:

- Universidades y entorno académico: fundamental desde los inicios, ya que fue en las universidades donde, por un lado, se comenzó a desarrollar toda la ciencia de datos y, por otro, algoritmos que hoy se tratan en el mundo empresarial como los últimos avances han sido conocidos en el ámbito científico desde los años sesenta.
- Incubadoras y aceleradoras de *startups*: muchas de las ideas que hoy surgen en torno a los datos son generadas por emprendedores que necesitan, por un lado, encontrar los socios o aliados estratégicos que los provean de la información suficiente y, por otro, de la financiación, la mentorización y la visión de negocio necesarias para sacar adelante una compañía. En ambas cuestiones, las incubadoras y aceleradoras de *startups* desempeñan un papel fundamental, pues, al eliminar barreras, los emprendedores se pueden focalizar en la propuesta de valor de su negocio.
- Regulador: el papel de los reguladores en el mundo de los datos es dual, ya que deben buscar el equilibrio entre la preservación de los derechos individuales de los usuarios y clientes, acotando los usos y comparticiones de datos entre empresas, y frenar con sus decisiones lo menos posible la capacidad de innovación, generación de valor y desarrollo de nuevas ideas por parte de empresas y emprendedores.
- Escuelas de formación: al ser la escasez de profesionales cualificados uno de los principales motivos de la deceleración en la implantación del *big data* y *data science* de forma masiva en las empresas, el papel de las escuelas de formación es esencial. Deben generar profesionales que sean capaces de tener cualidades en ciencias de la computación y bases de datos, matemáticas y estadística, visualización de datos y conocimiento del negocio concreto.

4.2

Ciclo de vida del dato: generación, captura, transmisión, almacenamiento, análisis y procesamiento

Como se ha visto anteriormente, la posición de cada empresa responde, de alguna manera, a un punto de la cadena de valor del dato: un registro se genera debido a una interacción concreta, se transmite a través de una red y se almacena en una infraestructura y, a partir de ahí, se integra con otras fuentes, se analiza y se finaliza construyendo otros servicios sobre estos datos o información derivada de ellos, pero ¿qué hace que cada uno de estos puntos tenga valor por sí mismo?, ¿por qué surgen empresas que se dedican a proporcionar estos servicios incluso, en ocasiones, de forma exclusiva? En el mundo del *big data*, cada aspecto de la cadena de valor del dato tiene sus complejidades técnicas y tecnológicas y, por ello, empresas de todo tipo han visto una oportunidad de negocio en ello.

4.2.1 Generación y captura de datos internos en las empresas

En esta sección se detallan algunas de las fuentes de datos más típicas internamente en las empresas. Debe pensarse que, para abordar todos los problemas que surgen en el día a día, las compañías utilizan varias de estas tecnologías que normalmente se encuentran separadas, por lo que la explotación de los datos que generan se convierte en algo nada trivial, incluso cuando en muchísimas ocasiones, grandes y pequeñas empresas por sí mismas no suponen una fuente de datos masivos.

CRM y herramientas de gestión comercial

CRM responde a las siglas de *customer relationship management*, herramientas que permiten a los departamentos comerciales de las empresas mantener la traza de su relación con los clientes. Por ejemplo, una empresa de consultoría almacena en estas herramientas información relacionada con clientes potenciales, información de empresas, identificación de personas clave, oportunidades comerciales, los estados por los que pasa, si se ganó o si se perdió, y todas las tareas que hubo que realizar para llegar a este desenlace. Disponer de toda esta traza permite obtener métricas del rendimiento de cada uno de los comerciales, de las zonas en que se alcanzó una mayor penetración, o del *funnel* (embudo) de conversión desde que se abre una oportunidad hasta que se materializa. Esta información es además la base para realizar modelos predictivos de *cross-selling* (si quiero vender otro producto a un cliente existente, ¿con quién debería contactar?), de *up-selling* (si tiene un producto, ¿a quién tengo que venderle la versión más avanzada?) o de captación (¿qué características tienen las empresas a las que consigo vender mi producto por primera vez?).

Bases de datos de procesos operativos y gestión de la productividad

Los procesos en las compañías, en ocasiones, constan de una cantidad enorme de pasos, involucran a muchísimos empleados, exigen muchas interacciones con el cliente final y dependen de otros objetos, como documentación, aceptaciones (firmas), etc. Cualquier mejora en la eficiencia de estos procesos haría que una determinada empresa, con los mismos recursos, fuera capaz de gestionar un mayor volumen de negocio. Imaginemos que una compañía de seguros de coche tiene una serie de personas dedicadas a gestionar los siniestros: un asegurado realiza un parte y abre una solicitud, un empleado verifica que toda la documentación es correcta y completa, se activa una petición a un perito para que analice el vehículo siniestrado, se recibe un informe y, tras todos estos pasos, se decide si se concede o no una prestación. Incluso en esta versión simplificada del proceso, una solicitud pasa por multitud de estados, cada uno con su rastro digital y el trabajo de diferentes empleados. Del análisis de toda esta traza pueden medirse tiempos de cada paso para detectar cuellos de botella o métricas de fraude en siniestros para cuantificar el dinero que se pierde por este motivo. Si gracias a estos datos, utilizando la estadística y los modelos predictivos, somos capaces de reducir tiempos o disminuir el fraude, estaremos dando un valor adicional a la compañía.

Logs y tag managers

Otra fuente de datos muy asociada al mundo de la Economía Digital está relacionada precisamente con la huella que dejamos cada vez que actuamos en un portal online: cada vez que visitamos una web, clicamos en una sección, en un producto, visualizamos un vídeo o realizamos una búsqueda, dejamos un registro, tanto de la interacción como de sus resultados. Esta traza queda almacenada en los llamados *logs* de los servidores web, las máquinas que son capaces de generar y mandar las webs que visitamos a nuestros navegadores. En ocasiones, las webs que visitamos son complejas y, para entender el comportamiento de los usuarios, es necesario monitorizar más hechos. Esto no se consigue solamente con los *logs* mencionados anteriormente sino utilizando unas herramientas que permiten etiquetar eventos y contenidos en la web, y monitorizarlos de manera exhaustiva, los denominados *tag managers*. Las empresas que utilizan este tipo de tecnologías son aquellas que tienen portales online y, en la actualidad, son prácticamente todas.

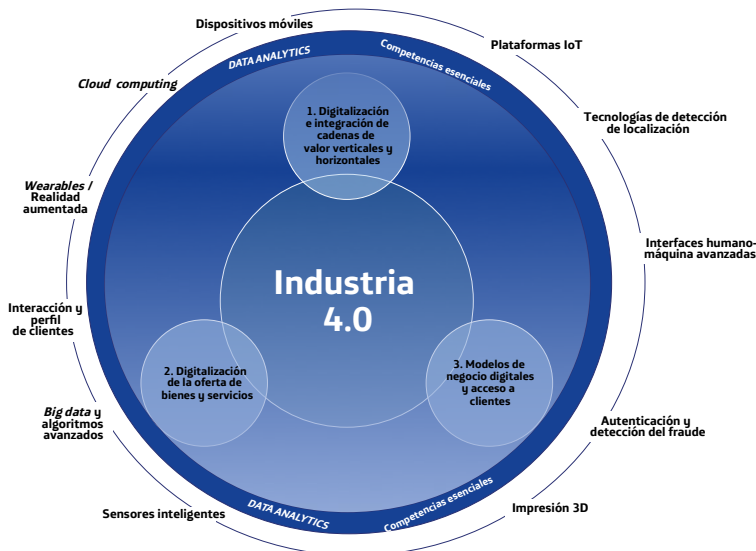
Interactuamos en el mundo online con medios de comunicación, compañías telefónicas, entidades financieras o tiendas de comercio electrónico y, dependiendo de los objetivos de cada una de ellas, mediremos unas situaciones u otras: por ejemplo, en comercio online, interesará tener muy detalladamente monitorizado todo lo que tenga que ver con interacciones con el «carrito de la compra», mientras que, en un negocio como un banco online, estará mucho más enfocado en los contenidos que visita, sobre lo que se informa o las secciones de atención al cliente.

Monitorización y sensorización: Industria 4.0

El concepto de industria 4.0 se refiere a la siguiente revolución industrial que consiste en la introducción de la digitalización en todos los procesos de fabricación, es decir, la utilización de tecnologías como el *big data*, la robotización, la simulación o la realidad aumentada dentro de las fábricas.

Fig. 4.1

Marco de la industria 4.0 y las tecnologías digitales



Fuente: Afi, a partir de datos de *Harvard Business Review*.

¿Por qué el *big data*? Por las múltiples fuentes de información derivadas de esta digitalización: por ejemplo, RFID es la identificación mediante tecnologías de radiofrecuencia de determinadas etiquetas. Es la misma tecnología que se utiliza en las tiendas de moda para detectar que alguien se ha llevado una prenda sin pagarla. En el contexto industrial, permite sensorizar a los empleados, detectando los lugares por donde van pasando y creando mapas de calor (qué zonas son más frecuentadas, cuáles son lugares de paso) o, por otro, paquetes y recursos físicos móviles en una fábrica.

Otro tipo de sensor es precisamente la localización por GPS que, en empresas con un fuerte componente logístico, permite tener controlada la flota y detectar ineficiencias en la gestión. También existen sensores para, por ejemplo, las líneas de ensamblaje, que facilitan controlar de manera detallada la producción de cada sección, detectar cuellos de botella y lanzar alertas automáticas ante determinados eventos. Por otro lado, herramientas como ERP (*Enterprise Resource Planning*) han

sido tradicionalmente utilizadas para gestionar la demanda, los pedidos salientes y las compras procedentes de los proveedores. Esta información, analizada correctamente y aplicando las técnicas de *data science* actuales, nos permite realizar predicciones sobre la demanda o estimar el riesgo económico de una empresa, en función de los proveedores y clientes que tiene.

Colaboración 1. El ciclo de vida del dato

Alberto Turégano, asesor de arquitecturas *business intelligence & big data*

En 1985, Michael Porter introdujo el concepto de cadena de valor empresarial —o simplemente cadena de valor— al modelo teórico que permite describir el desarrollo de las actividades llevadas a cabo por una empresa para generar valor al cliente final. Hoy en día, son muchas las empresas que siguen apoyándose en esta herramienta para diseñar su plan estratégico cuyo objetivo final es maximizar la creación de valor mientras se minimizan los costes. En otras palabras, el fin último es generar una ventaja competitiva.

De manera análoga, y debido a la reciente disrupción digital donde los datos se posicionan como un activo estratégico para las empresas, surge el concepto de cadena de valor del dato, donde se definen igualmente aquellas actividades y aquellos procesos que se deben llevar a cabo para poder monetizar el dato (ya sea de manera interna o bien externa) de la manera más eficiente, maximizando igualmente el valor para el cliente y generando una ventaja competitiva.

Existen múltiples versiones de la cadena de valor del dato, pero básicamente se podrían establecer las siguientes fases dentro de una cadena de valor estándar:

1. Fase de provisión del dato. Esta fase se descompondría a su vez en dos subfases:
 - 1.1 Aprovisionamiento-abastecimiento. Esta etapa se centra en identificar aquellos conjuntos de datos necesarios para dar soporte a un caso de uso específico. Uno de los grandes retos de las organizaciones es identificar qué fuentes de datos son las necesarias para pasar a formar parte de un análisis. Para ello, lo primero que habría que hacer sería realizar un inventario de aquella información que genera la empresa y que se está almacenando para, a continuación, realizar el inventario de aquella información que se está generando pero no se está almacenando y pudiera ser potencialmente útil para el caso de uso en cuestión. Finalmente, procederíamos a completar el inventario anterior con información externa a la organización, bien a través de terceros o bien incorporando información proveniente de fuentes abiertas (*open data*). Como aspecto clave, cabe señalar que cuanta más información se provisione a los científicos de datos, mejores serán los algoritmos que estos diseñen.

- 1.2** Captura y almacenamiento. En esta etapa se realizan las tareas necesarias para extraer la información de los sistemas fuente seleccionados en el paso anterior y para integrarlos en el sistema, listos para ser procesados y almacenados. Existe una gran diversidad de aproximaciones a la hora de capturar la información dependiendo fundamentalmente del tipo de los datos —estructurados, semiestructurados o desestructurados—, la frecuencia con la que se necesitan capturar —en tiempo real, *near real time* o *batch*—, y la exactitud y calidad de la misma. La buena noticia es que hoy en día disponemos de la tecnología necesaria para poder afrontar con garantías los retos anteriormente citados.
- 2.** Fase de transformación. Una vez que el dato se ha almacenado en el *data lake* (área de nuestro sistema *big data* donde se almacena toda la información, tanto la información en bruto como la ya disponible para su análisis y procesado), se procede a limpiarla y enriquecerla con el objetivo de proveer información precisa y de calidad (de confianza) a los usuarios, de tal forma que las decisiones basadas en dicha información se sustenten en información proveniente de fuentes fiables. El proceso para mejorar la calidad del dato es un proceso iterativo llamado *data wrangling* e incluye las siguientes tareas:
- 2.1** Perfilado. Consiste en realizar un análisis general de los datos (estadísticas) después de cada transformación, con el objetivo de determinar posibles incoherencias, de manera unitaria o respecto a la información histórica ya cargada previamente, de tal forma que se pudieran detectar anomalías en los datos.
- 2.2** Limpieza. Aquí se garantiza que cada registro del *dataset* (fichero o conjunto de datos) proporciona un valor válido en cada campo del mismo. Para ello se incluyen las siguientes tareas:
- Estandarización. Consiste en asegurar que datos del mismo tipo tienen el mismo formato, a fin de que luego puedan ser comparados o cruzados. Por ejemplo, los datos tipo fecha pueden venir con distinto formato de distintas fuentes. La solución sería estandarizar todas las fechas a un formato común, por ejemplo: dd/mm/aaaa.
 - Normalización. Consiste en la preparación de una lista de valores únicos que se pueden asignar a diferentes valores que representen la misma idea. Por ejemplo, crear una lista con los valores H y M para hombre y mujer, y asignarlos a todos los campos que recojan esa idea en cualquiera de las fuentes que estamos cargando. En el caso, por ejemplo, de que del sistema operacional de atención al cliente nos viniera el campo género con el valor varón, nuestro proceso de normalización lo transformaría al valor H.
 - *Matching*. Este proceso consiste en la eliminación de duplicados con el objetivo de mejorar la exactitud de la información de nuestro *data lake*.

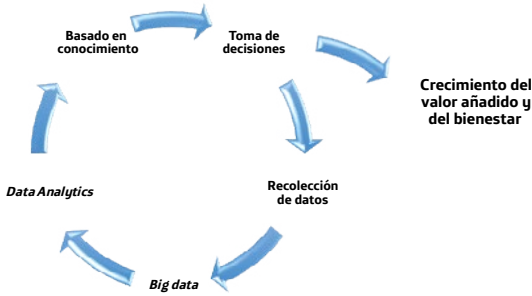
- 2.3** Estructuración. Si se considera necesario, se reestructuran los *datasets* de forma aislada. Se puede decidir dividir un campo como, por ejemplo, el que recoge información del nombre y apellidos de nuestros clientes en tres campos o viceversa.
- 2.4** Enriquecido. Se refiere a la tarea de enriquecer el *dataset* con nuevos campos que enriquezcan nuestro conjunto de datos. Por ejemplo, a la tabla de clientes podríamos añadirle campos con información relativa a la salida de diversos modelos analíticos, tales como la propensión a la compra de cierto producto o servicio o la probabilidad de baja (tasa de rotación).
- 3.** Fase de descubrimiento y modelado. Es en esta fase donde se realiza el proceso de análisis por parte de los analistas de negocio y los científicos de datos con el objetivo de extraer el máximo valor transformando los datos en información.
- 4.** Fase de exposición. Es la fase final de la cadena. En esta fase se realizan todas las tareas necesarias para poner el nuevo modelo en producción, puesto que no se genera valor si no se comparte el conocimiento extraído con los usuarios de negocio. Hay que tener en cuenta que en las fases anteriores se ha trabajado en modo «piloto», de tal forma que se han invertido los recursos mínimos necesarios para que en el caso de que no se extraiga valor de los *datasets* cargados en el *data lake*, la pérdida sea mínima también. Es, por tanto, en esta fase donde se realizan las tareas de validación de premisas asumidas anteriormente respecto a los datos de entrada, preparación del espacio necesario para albergar el modelo de salida, documentación del glosario de negocio con los nuevos términos o KPI, etc. En resumen, se industrializa todo el proceso anterior.

Todos los pasos detallados deben ser siempre ejecutados con el objetivo estratégico en mente, de manera que se gestione correctamente el alcance y el esfuerzo de cada fase y se centre en aquellas actividades que generen valor.

Por último, pero no por ello menos importante, es preciso señalar la idoneidad de establecer un programa de gobierno del dato que defina e implemente políticas, procedimientos, papeles y responsabilidades en cada una de las fases de la cadena de valor, con el objetivo de gestionar de manera eficiente los activos de información de la organización.

Fig. 4.2

El ciclo de vida del dato



Fuente: Afi.

4.2.2 Los datos de terceros como complemento

Las diferentes fuentes de datos internas en las empresas y algunas de las aplicaciones más directas que de ellas se derivan han sido descritas más arriba. Las empresas son, pues, generadoras de datos y los utilizan para los casos de uso que son capaces de definir con el objetivo de mejorar sus procesos y vender más.

Sin embargo, la visión de los datos de una compañía es siempre parcial: por ejemplo, un banco ve las cuentas que sus clientes tienen, pero no ve las cuentas que estos mismos clientes tienen en otros bancos; una compañía aérea sabe si alguien voló el año pasado con ellos, pero no si ese alguien lo hizo en otra compañía; una aseguradora puede saber cuánto vale el coche que conduce un cliente, pero no qué equipo de fútbol apoya. Esta pregunta, a priori, puede parecer que no tiene nada que ver con el negocio de los seguros, pero si el departamento de *marketing* de la aseguradora se está planteando un patrocinio deportivo, entonces, dicha pregunta, relevante, no puede responderse solo con los datos de los que dispone la empresa. Dependiendo de la pregunta que queramos resolver, los datos disponibles en la compañía pueden darnos una visión demasiado parcial que será necesario enriquecer con otras fuentes de información.

Como ya se ha comentado, una de las fuentes de datos que despierta más interés en las empresas son los datos de redes sociales, ya que allí se vierte información sobre el perfil (edad, ubicación, sexo, estado civil, etc.), opiniones, gustos, marcas preferidas, tendencias políticas, etc. Incluso un análisis pormenorizado puede ofrecer a las compañías información sobre cambios de domicilio, de situación familiar o laboral. De este modo, enriquecen de manera muy considerable la información que tienen internamente y que no podrían obtener de otra manera.

Una fuente de información externa aparentemente sencilla es la referida a los datos relacionados con calendarios y eventos, como los periodos vacacionales, los días festivos, los días en los que hay programados determinados conciertos, ma-

nifestaciones, etc. Si un comercio online quiere predecir la demanda que va a tener de sus artículos, los calendarios y eventos influyen de manera decisiva en la predicción sobre el número de pedidos que habrán de ser gestionados. También se producen fenómenos estacionales nuevos que hacen crecer el número de pedidos de manera exponencial: si los modelos de previsión de demanda no tuvieran en cuenta fechas como el *Black Friday*, fallarían estrepitosamente en una de las épocas del año donde más compras online se realizan. En este mismo contexto, los datos sobre predicciones climatológicas pueden indicar que los usuarios no saldrán a la calle y que, gracias a las facilidades de la tecnología, realizarán sus compras a través de portales online y, por tanto, la demanda por este canal se incrementará, con toda la producción logística que ello implica.

Finalmente, los Gobiernos y las Administraciones públicas son una parte fundamental en este ecosistema de datos de terceros. En España existen portales como el del Instituto Nacional de Estadística (INE) que cuentan con un repositorio de datos sobre sociodemografía, padrón o actividad económica a diferentes niveles geográficos y/o administrativos. Esta información puede ser aprovechada de manera sustancial en aplicaciones relacionadas con la educación o, simplemente, como base para un estudio de mercado que permita tomar decisiones comerciales.

Desde hace unos años existe también un Portal de Transparencia (www.transparencia.gob.es), donde se puede encontrar toda la información pública acerca de presupuestos de los ministerios, adjudicaciones de contratos o políticos. Y es que los datos no solamente tienen una fuerte capacidad de hacer crecer nuestra economía, sino también la participación e información ciudadana.



4.3

¿Dónde se almacena el *big data*?

Aunque para sacar valor de los datos no hace falta tener volúmenes que las máquinas tradicionales no puedan procesar (el término *big data*, en ocasiones, ha generado confusión a este respecto), cada día más empresas sí que enfrentan un problema de volumen de datos y su almacenamiento. Por ejemplo, las compañías que se dedican a la gestión del *marketing* digital de sus clientes, invirtiendo el dinero de las campañas publicitarias de manera inteligente para maximizar sus resultados, necesitan la traza de por dónde pasa el usuario, dónde ve la publicidad o si hace clic o no. Si una de estas compañías tiene como cliente a cinco empresas de tamaño medio (de unos cincuenta mil clientes), puesto que está midiendo de manera continua métricas de éxito de sus campañas y registrando cada interacción y cada conversión de cada usuario, puede suceder que en relativo poco tiempo, esta compañía tenga que gestionar cientos o incluso miles de millones de registros, lo que sí se convierte en un problema.

Tradicionalmente, las empresas han optado por una aproximación que típicamente se ha acordado en llamar *data warehouse*, grandes bases de datos estructuradas donde se ha intentado programar toda la lógica operativa de la empresa en tablas y con una rigidez fuerte en la información que la contiene. Además, estas bases de datos solían estar almacenadas en máquinas que albergaban las empresas en sus centros de procesamiento de datos (CPD) y que, por tanto, requerían un mantenimiento. Esto implicaba que cada cierto número de años fuera necesario cambiar la tecnología, ya que la evolución de la misma en los últimos tiempos ha sido vertiginosa. Si al hecho de que estas tecnologías no son las mejores para almacenar grandes volúmenes de información, se une el que no escalan de manera sencilla y que, además, suelen ser caras, no parecen ser las tecnologías óptimas para abordar el problema del volumen en los proyectos *big data*.

Al facilitar el procesado de grandes tablas de datos transaccionales con miles de millones de registros, la aparición de tecnologías como Hadoop (véase el capítulo 2) permitió escalar las bases de datos de las empresas de una manera mucho más sencilla, pero la sola adopción de esta tecnología no resuelve el hecho de que la inversión en máquinas capaces de llevar a cabo toda esta operativa sea una decisión cara que puede alejar a las pymes del apasionante mundo de los datos.

Paralelamente a la aparición de las tecnologías *big data*, otra de las tendencias ha sido la adopción del *cloud computing* o computación en la nube, por la cual las empresas pueden liberarse de la compra de máquinas, su mantenimiento y, en definitiva, del coste en recursos que supone la operativa de una gran infraestructura informática. Determinadas tecnologías de *cloud computing* permiten a las empresas escalar sus infraestructuras bajo demanda, esto es, pedir más capacidad de almacenamiento y procesamiento según su necesidad, pagando solamente por el

tiempo que consumen estas capacidades. Esto permite, por ejemplo, levantar un grupo (*cluster*) de veinte servidores de manera muy rápida, mantenerlo durante el cálculo de una serie de métricas y, una vez terminado el cálculo, destruir el *cluster*. Los proveedores de infraestructura como servicios (IaaS son sus siglas en inglés) facturan a sus clientes solamente por el tiempo de uso de la infraestructura levantada para los cálculos, lo que hace que analizar los datos a gran escala se haya convertido en algo accesible para todo tipo y tamaño de compañía de una manera ágil, barata y eficaz.

4.4

El análisis de datos y sus profesionales: estado actual, demanda actual, *gap* y previsiones futuras

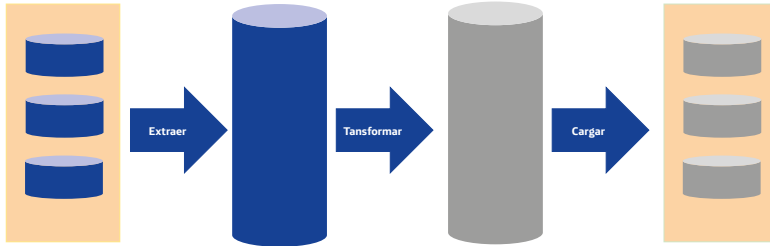
Según las estadísticas elaboradas en la Unión Europea, en 2016 existían unos 350.000 empleos relacionados con los datos en España y en torno a 6 millones en toda la Unión Europea. Aun con estas cifras, la demanda de este tipo de profesionales en la Unión Europea roza en la actualidad el medio millón de puestos de trabajo, con previsiones de alcanzar una escasez de 800.000 puestos en 2020. ¿Por qué sucede esto?

Dentro del mundo *big data*, existen diferentes puestos profesionales que dependen de sus tareas en el ciclo de datos analizado. En general, se suele hacer distinción entre dos perfiles: por un lado, los ingenieros de datos, responsables de los procesos de ingesta de datos en las infraestructuras y de la propia operativa de las mismas y, por otro, los *data scientists*, responsables de analizarlos, crear modelos estadísticos y predictivos que permitan anticipar comportamientos y extraer el valor de los datos.

Las aptitudes necesarias en los ingenieros de datos son de perfil puramente técnico, y deben ser profundos conocedores de las tecnologías de almacenamiento, captura y procesamiento del *big data*. Esto implica el conocimiento de lenguajes de programación adecuados y con funcionalidades de *big data*, herramientas que ayudan a realizar procesos de migración de datos y transformación (procesos *extract-transform-load*, ETL), así como herramientas de extracción de datos, una vez están en este tipo de infraestructuras. Todo ello tiene poca relación con el *core* de negocio de las empresas al tiempo que demandan un componente técnico fundamental. Aunque estas tecnologías son relativamente modernas, la adaptación de profesionales del mundo del desarrollo de aplicaciones o la gestión y la utilización de bases de datos tradicionales se puede realizar de manera ágil, adquiriendo el conocimiento sobre estas tecnologías.

Fig. 4.3

Proceso ETL: extraer-transformar-cargar



Fuente: Afi.

En el perfil más analítico —el científico de datos— se mezclan distintos tipos de conocimiento. Para utilizar las infraestructuras donde se almacena el *big data* es necesario conocer lenguajes de programación y de extracción de datos, aptitud que se adquiere en carreras como Ingeniería Informática o de Telecomunicaciones. Por otro lado, es necesario el conocimiento de las herramientas estadísticas y matemáticas que permitan encontrar patrones en los datos o generar modelos predictivos, habilidades que han estado tradicionalmente repartidas entre diferentes carreras. Además de estos conocimientos técnicos, se suele requerir conocimiento del negocio, capacidades de comunicación y de gran entendimiento del cliente, lo que hace que el perfil de estos profesionales sea muy escaso, fundamentalmente por dos razones: en primer lugar, porque tradicionalmente no ha habido carreras universitarias que abordan de manera unificada todas estas actitudes y, aunque ahora sí existen, las nuevas generaciones de profesionales todavía se están formando o tienen poca experiencia; en segundo lugar, contrariamente al perfil de ingeniero de datos que, aunque costoso, puede tener una fuente de profesionales reconvertidos de otros puestos relacionados con la gestión de bases de datos, es mucho más difícil formar a profesionales ya cualificados en algunos de los aspectos, ya que los requisitos son de naturaleza muy diferente.

Fig. 4.4

Ciencia de datos: el trabajo del siglo XXI

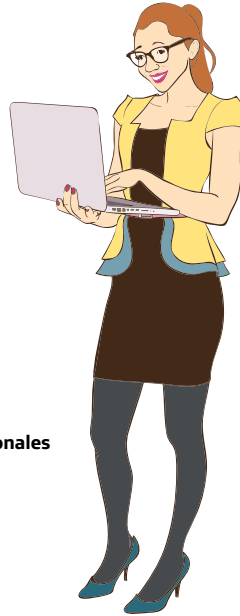
La ciencia de datos requiere una combinación de habilidades multidisciplinares, en la intersección entre las matemáticas, estadística, informática, comunicación y negocio.

Matemáticas y estadística

- *Machine learning*.
- Modelación estadística.
- Diseño experimental.
- Inferencia bayesiana.
- Aprendizaje supervisado: árboles de decisión, regresión logística, bosques aleatorios.
- Aprendizaje no supervisado: algoritmos de agrupamiento, reducción de dimensionalidad.
- Optimización: *gradient descent* y variaciones.

Conocimiento del negocio y habilidades emocionales e interpersonales

- Pasión por el negocio.
- Curiosidad sobre los datos.
- Capacidad de influencia sin autoridad.
- Mentalidad *hacker*.
- Solucionador de problemas.
- Estrategia, proactividad, creatividad, innovación y colaboración.



Programación y bases de datos

- Fundamentos de ciencias de la computación.
- Lenguaje (p.e. Python).
- Paquetes de computación estadística (p.e. R).
- Bases de datos: SQL, NoSQL.
- Álgebra relacional.
- Bases de datos paralelas y procesamiento paralelo de consultas.
- Conceptos *MapReduce*.
- Hadoop y Hive/Pig.
- *Custom reducers*.
- Experiencia con xaaS como AWS.

Comunicación y visualización

- Capacidad para interactuar con alta gerencia.
- Habilidad narrativa.
- Traducción de aprendizajes basados en datos a decisiones y acciones.
- Diseño artístico visual.
- Paquetes R como *ggplot* o *lattice*.
- Conocimiento de herramientas de visualización (p.e. Flare, D2.js, Tableau).

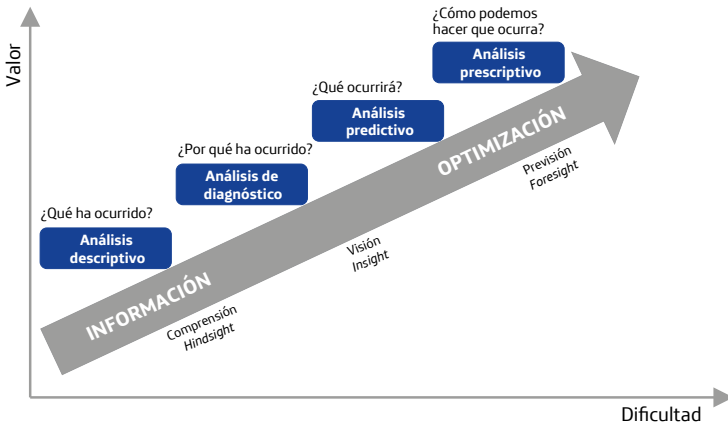
Fuente: The Modern Data Scientist, Marketing Distillery Blog, <http://www.marketingdistillery.com/>

Dentro de las actividades que llevan a cabo los *data scientists*, se engloban las siguientes:

- **Análisis descriptivo:** la aplicación de técnicas estadísticas que permiten extraer características generales de un conjunto de datos.
- **Análisis predictivo:** la utilización de técnicas procedentes del aprendizaje automático (una rama de la inteligencia artificial) para, a partir de los patrones históricos de los datos, calcular de manera automática por estos algoritmos una serie de modelos matemáticos que, ante nuevos datos, son capaces de pronosticar el futuro.
- **Análisis prescriptivo:** en muchas ocasiones no es suficiente la sola aportación de predicciones, pues, para implementar un sistema completamente inteligente, los algoritmos deben proporcionar, además, la acción recomendada.

Fig. 4.5

Del análisis descriptivo al prescriptivo

Fuente: <http://www.rosebt.com/blog/descriptive-diagnostic-predictive-prescriptive-analytics>

Recurriendo a un ejemplo de un comercio online: a partir de los datos de navegación de la compañía, se pueden detectar puntos de fuga en el proceso de compra y, con ello, tomar decisiones sobre el diseño de la página web (análisis descriptivo); en otra fase del proyecto puede generarse una puntuación sobre la probabilidad de que un cliente cause baja en el servicio (análisis predictivo). Si además otro algoritmo dice el canal a través del cual debe contactarse con los clientes que van a abandonar (email, teléfono, visita presencial, oferta, etc.) para retenerlos con una mayor probabilidad de éxito, el círculo de la toma de decisiones de manera automática se cerrará (análisis prescriptivo).

Puesto que para llevar a cabo este tipo de análisis se requieren aptitudes diversas dentro de competencias informáticas, matemáticas y de conocimiento del negocio, hoy en día existe una necesidad muy alta de formación especializada, al menos, en tres vertientes diferentes:

- Formación que genere científicos de datos profesionales.
- Formación de *data scientist* líderes: profesionales que, aunque no precisen conocer el detalle técnico de los algoritmos, sean conscientes de su complejidad, de las métricas que se pueden calcular y de su rendimiento con la suficiente experiencia para aplicar sus resultados al negocio.
- Formación y cultura del dato para ejecutivos: la utilización de los datos es un cambio cultural en las empresas que modifica la forma de abordar los problemas y que mueve un paradigma de toma de decisiones basadas en el conocimiento experto más racional, a partir de la evidencia estadística. Es por ello que, aunque no sean los responsables de los equipos de *data science*, sí deben de ser capaces de definir sus necesidades y hacerlo en términos de datos, analítica y predicciones.

4.5

Evolución del volumen de datos y su utilización en la Economía Digital

La Unión Europea, en diversos estudios publicados en 2016,¹ ha estimado el valor que la Economía de los Datos está aportando en los últimos años, y ha proyectado cuál será su valor para los próximos con el horizonte puesto en 2020. El valor de la Economía de los Datos en Europa alcanzó los 285.000 millones de euros en 2015, representando así algo menos del 2% del total del PIB de la Unión Europea. El mismo organismo estima que, si las condiciones legislativas son propicias y son definidas a tiempo, en 2020 el valor de la Economía de los Datos alcanzará los 739.000 millones, esto es, un 4% del total del PIB europeo.

El mismo estudio ha encontrado que unos 6,2 millones de personas en Europa trabajan en empleos relacionados con los datos. En los escenarios más optimistas, a medio plazo, esta cifra alcanzaría prácticamente 10,5 millones de empleos, con unas tasas de crecimiento superiores al 14%. Todo este mercado involucrará también a más de 359.000 empresas en 2020, aproximadamente un 9% del total de las mismas.

Otro de los aspectos fundamentales es la aparición de empresas que crean productos y servicios basados en datos. Según la Unión Europea, en 2016 todas estas empresas generaron más de 59.000 millones de euros y las previsiones para los años venideros atisban un incremento de alrededor del 15% en esta cifra de negocio.

Por sectores, en toda Europa, los que más están apostando por el *big data* son las manufacturas y las finanzas y, según todas las previsiones, seguirán estando a la cabeza en lo que a implantación de estas tecnologías supone.

Tabla 4.1 Sectores de actividad con mayor relevancia en *big data*

	Manufacturas	Finanzas	Servicios profesionales	Retail	Media
Valor de mercado 2016	12,8 K M€	11,8 K M€	8,4 K M€	6,3 K M€	5,8 K M€
Previsiones 2020	17,3 K M€	15,6 K M€	11,3 K M€	8,3 K M€	8,2 K M€

Fuente: Afi a partir de datos de la Comisión Europea.

En España, esta tendencia también se observa en lo que respecta a los diferentes actores del mercado de los datos. Se estima que la Economía de los Datos supuso en España una generación de valor (negocio) de más de 3.200 millones de euros

1. <https://ec.europa.eu/digital-single-market/en/policies/building-european-data-economy>, <https://ec.europa.eu/digital-single-market/en/towards-thriving-data-driven-economy>, <http://www.datalandscape.eu/>

en 2016, con una perspectiva de alcanzar los 4.000 millones de euros en 2020. En los otros aspectos existen tendencias de crecimiento similares a las del resto de la Unión Europea, lo que hace prever que el mercado de los datos en España crezca de manera muy significativa en los próximos años.

Colaboración 2. El valor de los datos para las instituciones públicas

Alejandro Llorente, científico de datos y cofundador de PiperLab, S.L.

Las instituciones públicas desempeñan un papel fundamental en la Economía de los Datos. Todas ellas son grandes generadoras de datos: realizan encuestas para contabilizar censos y extraer estadísticas sociodemográficas de la población, sensorizan ciudades para optimizar las políticas que se aplican, elaboran bases de datos con presupuestos públicos, publican cientos de contratos públicos, reciben toda nuestra información fiscal, tanto de personas físicas como de empresas y, así, un sinfín de datos que permiten a las Administraciones de todos los niveles funcionar en su día a día.

Pero, aparte de generar y recolectar datos, ¿cómo impacta el *big data* y la *data science* en las Administraciones y las instituciones públicas? Más allá de ofrecer los datos recolectados como servicio y ejercicio de transparencia, ¿pueden los datos optimizar la gestión pública y ayudar a los políticos y funcionarios a tomar mejores decisiones?

Uno de los principales retos a los que se enfrentan los Gobiernos es la gestión de los efectos del cambio climático y el diseño de acciones preventivas que mitiguen los posibles sufrimientos asociados. En este sentido, existen entidades, como el Servicio Nacional de Tiempo de Estados Unidos, que recolectan terabytes de datos meteorológicos alrededor del mundo. Esta información permite a los investigadores generar modelos para comprender mejor el cambio climático, estimar tendencias y generar predicciones a largo plazo sobre la frecuencia de huracanes, inundaciones o fuertes sequías. Con todos estos estudios, los Gobiernos pueden predecir, por ejemplo, la disponibilidad futura de agua potable y, junto a predicciones de demanda derivadas de proyecciones demográficas, tomar decisiones sobre su gestión.

Otra de las áreas en el contexto público es la aplicación del *big data* y de la *data science* en servicios sociales. Toda la información sobre impuestos, peticiones de subsidios y ayudas, así como estadísticas públicas, permiten a las Administraciones optimizar sus procesos de gestión, adjudicación de ayudas de manera mucho más rápida y eficaz. Además, con esta información se puede, entre otras cosas, detectar patrones de comportamiento y predecir y prevenir fraudes de distintos tipos, como en la gestión de bajas médicas o en las peticiones de subsidios, donde se modifica la realidad para cumplir con los requerimientos.

En el ámbito de la gestión médica, se almacenan enormes volúmenes de información procedente de hospitales, centros de salud y otros servicios. Por ejemplo, todas las historias clínicas, informes de accidentes, actuaciones de urgencia y ambulancias y ficheros de servicios sociales o información generada por los centros de control de epidemias son algunas de las fuentes de este sector. Si se cruza toda esta información con la información sociodemográfica de los ámbitos de influencia de los mismos, se puede estimar la demanda de distintos tipos de servicios médicos y detectar de manera temprana determinadas amenazas sanitarias dependiendo de las zonas. En la actualidad, se utilizan todos estos datos y otros adicionales, como las búsquedas en Google o datos de redes sociales, para monitorizar y anticipar epidemias estacionales que saturan los servicios de urgencia, como la gripe invernal o determinados virus estomacales. Si los gestores sanitarios pueden anticipar qué servicios se van a necesitar, podrán optimizar los recursos que hay en ellos y, por tanto, aumentar la calidad del servicio que todos recibimos.

Por último, otro de los sectores de la Administración pública donde se utilizan ampliamente las técnicas de análisis estadístico de datos es en el ámbito de la seguridad. Con información integrada procedente de diferentes fuentes de información —comunicaciones, movimientos, llamadas o redes sociales—, el estudio de los patrones de comportamiento y de las anomalías que en estos datos se detectan permite la identificación de amenazas potenciales en temas tan diferentes como terrorismo doméstico y/o internacional, análisis de las causas de la delincuencia y anticipación de crímenes, ayuda hoy vital en la lucha contra el ciberterrorismo.

En definitiva, los datos están cambiando la forma en la que todos tomamos nuestras decisiones: de la misma manera que las aplicaciones digitales nos ayudan a monitorizar nuestra salud, recomendar acciones basadas en nuestros comportamientos y compras pasadas en cualquier comercio online, las fuentes de datos —tanto generadas por la Administración pública en su propia operativa, como las disponibles de manera externa (datos de terceros)— permiten a las personas que toman decisiones de gestión y política pública disponer de una mayor información. Esto supone avances como la previsión y la optimización de recursos en ámbitos tan distintos como la sanidad, la seguridad o la prevención de fraude a la Administración. Además, cada vez existen más iniciativas por parte de las instituciones públicas de abrir sus datos a la comunidad de científicos de datos, a periodistas o, en general, a cualquier persona que tenga interés en ellos, en un ejercicio de transparencia que acerca la gestión a la ciudadanía. En España, disponemos de varios portales de datos abiertos, tanto a nivel autonómico como a nivel estatal, que pueden ser utilizados para conocer mejor el funcionamiento de nuestras instituciones.

4.6

Open data: concepto, limitaciones y medidas adoptadas en favor de la creación de estándares

*Open data*² es el término que se emplea para referirse a aquellos datos cuyo acceso es libre (es decir, no existen limitaciones técnicas, ni legales), y pueden ser usados, reutilizados y redistribuidos por cualquier usuario que lo desee.

Una definición más amplia y canónica del término incluye tres características que debe cumplir y que le son intrínsecas:

- Disponibilidad y acceso: la información debe estar disponible y debe poder ser extraída con facilidad e, idealmente, sin un coste adicional que el de la propia conexión a Internet que utilice el consumidor de la información.
- Reutilización y redistribución: los datos deben admitir la posibilidad de ser reutilizados y redistribuidos, así como ser fusionados con otras bases de datos.
- Participación universal: cualquier individuo, sociedad o institución debe poder tener acceso, reutilizar o redistribuir el contenido, es decir, no debe existir discriminación hacia ningún grupo, individuo o campo de actividad.

Es una práctica que persigue que determinados tipos de datos estén disponibles de forma libre para todo el mundo, sin restricciones ni mecanismos de control. Wikipedia, que opera bajo este concepto, expone que «tiene una ética similar a otros movimientos y comunidades abiertos, como el *software* libre, el código abierto (*open source*, en inglés) y el acceso libre (*open access*, en inglés).

Son considerados datos abiertos todos aquellos datos accesibles y reutilizables, sin exigencia de permisos específicos, aunque los tipos de reutilización pueden estar controlados mediante algún tipo de licencia. Se trata de fuentes de datos que históricamente han estado bajo el control de organizaciones —públicas o privadas— y cuyo acceso ha estado restringido. Estas limitaciones de acceso y uso van en contra del bien común, dado que se trata de información que pertenece a la sociedad o que ha sido generada u obtenida, en muchos casos, por Administraciones públicas financiadas por la ciudadanía.

2. Open Data Handbook Documentation, Release 1.0.0.

Fig. 4.6

Los ocho principios del *open data*

Fuente: Open Government Working Group (Estados Unidos).

La apertura de datos fomenta tanto la transparencia como la construcción de nuevos modelos de negocio sustentados en el análisis y generación de valor a partir de dichos datos, con el *big data* y el análisis científico de datos como herramientas de trabajo. La aplicación de técnicas de *social business analytics* sobre los datos capturados busca incrementar la atención de los usuarios, fidelizarlos y optimizar las campañas informativas, de sensibilización o participación realizadas, recopilar patrones de comportamiento y opiniones o preocupaciones en redes sociales para, en definitiva, diseñar e implementar estrategias acordes a la información recabada. Las fuentes de datos heterogéneas, las técnicas de *big data* y los algoritmos inteligentes permiten conocer diferentes perspectivas y diseñar soluciones para mejorar la vida en las ciudades.

En España, ha sido el sector público el que ha acometido en primera instancia la apertura de datos y todas aquellas Administraciones que aún no lo han hecho se verán pronto obligadas a implementar lo prescrito por la nueva Directiva 2013/37/EU sobre reutilización de información del sector público.³ Este hecho, además de pro-

3. La Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público es la norma que transpone al ordenamiento español la Directiva Europea 2003/98/CE que contiene la regulación básica del régimen jurídico aplicable a la reutilización de la información elaborada o custodiada por las Administraciones y los organismos del sector público. Esta ley se desarrolla para el ámbito público estatal a través del Real Decreto 1495/2011, de 24 de octubre, que introduce el derecho a la reutilización de cuanto documento tenga carácter público, recogido posteriormente en el marco de la Unión en la Directiva 2013/37/UE del Parlamento Europeo y del Consejo, de 26 de junio de 2013, por la que se modifica la Directiva 2003/98/CE relativa a la reutilización de la información del sector público.

mover la transparencia en la acción gubernamental, permitirá que cualquier agente con capacidades analíticas comprenda mejor el funcionamiento de las instituciones públicas y el buen uso de los recursos económicos.

Otras recomendaciones a nivel nacional e internacional que se deben considerar son las siguientes:

- La norma técnica nacional de interoperabilidad de reutilización de recursos de información.
- Las directrices del World Wide Web Consortium (W3C) sobre la publicación abierta de datos de las Administraciones.
- Las pautas para políticas *open data* definidas por la Sunlight Foundation.
- La declaración sobre *open data* realizada por la Global Open Data Initiative desde la perspectiva de la sociedad civil.
- Los principios y compromisos *open data* definidos por el G8.
- Las pautas de Naciones Unidas para la participación ciudadana en el *Open Government Data*.
- Las normas para ciudades inteligentes en desarrollo dentro del comité técnico AEN/CTN 178 de AENOR y, en particular, con la guía de especificaciones de datos abiertos prevista para el primer trimestre de 2015 dentro del trabajo del subcomité de gobierno y movilidad.

Los proyectos *open data* son un ejemplo de iniciativas en las que los órganos gestores de dichos datos no cuentan por sí mismos con todos los recursos para abordar el cambio necesario y requieren construir un ecosistema activo y equilibrado de usuarios para funcionar adecuadamente, esto es, la comunidad del *open data*.

Colaboración 3. El valor del *open data* en América Latina

Carla M. Bonina, Surrey Centre for the Digital Economy, Surrey Business School

El entusiasmo por los beneficios económicos y sociales que conlleva la utilización de datos en formato abierto, los llamados datos abiertos u *open data*, ha llegado para quedarse. En los últimos cinco años, Gobiernos de todo el mundo han incorporado progresivamente, una agenda de gobierno abierto, siguiendo a países como Reino Unido y Estados Unidos, los primeros en implementarla. América Latina no es la excepción. México, Brasil y Uruguay se encuentran entre los veinte primeros países del Barómetro de Datos Abiertos 2015 (2015 Open Data Barometer), un *ranking* mundial liderado por la World Wide Web Foundation, que califica a los países según su nivel de publicación de datos gubernamentales clave, su disposición para sacar el máximo provecho de los datos abiertos y las pruebas del impacto generado por los mismos. Colombia, por su parte, ocupó el cuarto lugar en el Índice de Datos Abiertos en 2015, una encuesta coordinada por la Open Knowledge Foundation que mide el estado de los datos de gobierno abierto en todo el mundo.

El valor de los datos abiertos

A medida que las empresas, los Gobiernos y los usuarios finales utilizan o reutilizan los datos abiertos, surgen nuevos productos, procesos y oportunidades de negocio. En 2011, un estudio encargado por la Comisión Europea estimó que el valor económico de divulgar y reutilizar la información del sector público rondaba los 40.000 millones de euros anuales solo en la Unión Europea.⁴ A escala mundial, McKinsey estimó en 2013 que el valor potencial de los datos abiertos llegaría a estar entre uno y tres billones de dólares por año.⁵ Si bien estas cifras no dejan de ser especulativas, la lógica de que los datos abiertos contribuyen a generar beneficio económico proviene de varios frentes. Para los Gobiernos, los datos abiertos pueden mejorar la prestación de servicios públicos a través de la modernización de los mismos: ayudan a reducir costes al evitar la duplicidad en la recogida de datos, y contribuyen a una mayor eficiencia en las operaciones internas. Para las empresas, estos prometen oportunidades para mejorar la toma de decisiones estratégicas, estimular la innovación e incrementar la productividad.

Por otro lado, los datos abiertos están ayudando a crear una nueva industria de datos en sí misma. Han surgido nuevas empresas que capturan valor recopilando, analizando y combinando los datos abiertos de distintas fuentes, u ofreciendo un servicio directo basado en datos abiertos. The Climate Corporation, en Estados Unidos, por ejemplo, constituye un caso notable de una empresa que construyó su *business case* utilizando datos meteorológicos que posteriormente fueron comprados por Monsanto por 930 millones de dólares en 2009.

Asimismo, cabe destacar que los datos abiertos se consideran especialmente beneficiosos para aumentar la transparencia, luchar contra la corrupción y promover la inclusión y el desarrollo socioeconómico, a pesar de que el valor de dichos beneficios sea mucho más difícil de evaluar en términos monetarios. De hecho, las iniciativas de implementar el uso de datos abiertos en los Gobiernos líderes (en particular, Reino Unido y Estados Unidos) surgieron como resultado de la falta de confianza en el sistema político, y de una puja por promover mayor transparencia en los Estados democráticos. Esta perspectiva no centrada en el valor económico de los datos ha sido especialmente importante en América Latina, donde la falta de rendición de cuentas y transparencia, junto con las necesidades sociales, ha estado en el centro de la agenda social y política.

El marco de los datos abiertos en América Latina: un ecosistema activo y en crecimiento

A pesar del aclamado potencial económico, el movimiento de datos abiertos en América Latina se puede entender mejor por su capacidad para generar impacto social y democrático, que por crear nuevos negocios con fines lucrativos. Dicho esto,

4. Vickery, G., «Review of Recent PSI Re-Use Studies». Bruselas: European Commission Directorate General/Information Society, 2011.

5. Manjika, J., Chui, M., Farrell, D., Kuiken, S.V., Groves, P. y Doshi, E.A., «Open Data: Unlocking Innovation and Performance with Liquid Information», McKinsey Global Institute, 2013.

Junar es quizá una de las pocas excepciones en la región.⁶ Junar, una empresa latinoamericana con oficinas en Silicon Valley, encontró un nicho de mercado al proporcionar una plataforma en la nube para Gobiernos y empresas que deseen publicar su información en formato abierto. Los clientes de Junar incluyen la ciudad estadounidense de Sacramento, la ciudad argentina de Buenos Aires y el Gobierno de Costa Rica. En comparación, otros países, como es el caso de Reino Unido, han establecido un modo de crear negocio a partir de los datos abiertos. Por ejemplo, el Open Data Institute (ODI), una compañía sin ánimo de lucro financiada por el Gobierno británico, con una red mundial en expansión, se fundó en 2012 con el objetivo de ampliar la red de negocio, y ha ayudado a diferentes *startups* a abrirse camino en el sector. El programa destinado a *startups* de ODI afirma que para 2015 había generado más de nueve millones de libras esterlinas a través de su red.⁷

En América Latina, el enfoque comercial de los datos abiertos aún está por despegar. Por el contrario, lo que actualmente es más bien único es la movilización de los actores que han impulsado la agenda de los datos abiertos. El ecosistema de los datos abiertos en América Latina se beneficia de un movimiento social que sigue creciendo y que incluye a los defensores de la transparencia y los datos, los Gobiernos y los desarrolladores, las agencias desarrolladoras de *software* y los donantes internacionales. La región inició su primera Conferencia de Datos Abiertos (AbreLatam y ConDatos) en Montevideo, Uruguay, en 2013, y se reúne anualmente desde entonces. Este esfuerzo pionero fortaleció a las comunidades regionales, defendiendo y utilizando los datos abiertos para aportar soluciones tangibles a los servicios públicos, y promovió la inclusión y el desarrollo social y económico. Según una encuesta realizada en AbreLatam, existen 196 proyectos de 135 organizaciones en la región, que abordan desde transparencia y periodismo hasta servicios de salud y aplicaciones relacionadas con la ciudad.⁸ Hoy la región cuenta con el centro de investigación Iniciativa Latinoamericana de Datos Abiertos (ILDA), que trabaja para poner a prueba y validar soluciones basadas en datos abiertos en la región.⁹

El ecosistema en América Latina significa que diferentes actores han contribuido a poner en valor los datos abiertos. De hecho, los Gobiernos son actores importantes en este ecosistema, ya que representan el lado de la oferta en la cadena de valor de los datos. Ciudades como Buenos Aires, Ciudad de México y Montevideo tienen plataformas de datos abiertos que son un referente en la región. El valor de estas plataformas reside en las aplicaciones y funciones que se puedan desarrollar a partir de estos datos. En la ciudad de Buenos Aires, por ejemplo, se sirvieron de los encuentros *hackathons* y otros concursos de aplicaciones para invitar a desarrolladores y *startups* a colaborar en la cocreación de nuevos servicios. En Montevideo, el municipio se asoció con una organización de sociedad civil de datos abiertos llamada

6. Website: <http://junar.com/>

7. Fuente: <https://theodi.org/start-ups>

8. <http://abrelatam.org/>

9. <http://www.idatosabiertos.org/>

Datos Abiertos, Transparencia y Acceso a la Información (DATA) para desarrollar la plataforma Por Mi Barrio. Inspirada en el sitio web FixMyStreet (ArreglaMyCalle) de Reino Unido, Por Mi Barrio ayuda a los ciudadanos a informar de problemas en la vía pública, como farolas rotas o baches, y los remite a la autoridad municipal correspondiente para su solución.

Las ONG y los activistas locales son, por consiguiente, otro actor crucial en este ecosistema. Nuevamente en Uruguay, DATA dirigió varios proyectos que terminaron transformando el modo en que operan y cooperan el Gobierno y los ciudadanos. Utilizando los datos de gobierno abierto en materia de salud, se asociaron con el Ministerio de Salud, y lanzaron A Tu Servicio, un sitio web intuitivo que ayuda a los uruguayos a comparar proveedores de salud locales, y estar mejor informados en la toma de decisiones.¹⁰ Otros casos similares son Ciudadano Inteligente en Chile y SocialTIC en México. Ambas organizaciones promueven la transparencia, la inclusión y la participación ciudadana a través de datos abiertos y el uso de nuevas tecnologías de la información.

El periodismo, por su parte, también ha contribuido a que los Gobiernos publiquen sus datos en formato abierto con el fin de rendir cuentas. En la región, un caso destacable es el de *La Nación* y su equipo de datos (LN Data).¹¹ Este periódico argentino realiza trabajo innovador con los datos abiertos, y ha sido reconocido internacionalmente con varios premios en su sector. En 2013 ganó el Premio de Periodismo de Datos por su trabajo en la apertura de los datos no estructurados, cerrados y opacos de los gastos del Senado en Argentina, un país sin leyes que regulen la libertad de información y con un historial de restricciones de acceso de los medios a la información gubernamental. En Perú, la organización Convoca liberó datos públicos para ayudar a los usuarios a comprender el comportamiento de las industrias extractivas en el país y su impacto en la vida de los ciudadanos. Este proyecto, llamado Excesos sin Castigo, ganó el Premio de Periodismo de Datos en 2016.¹² Poderopedia, una plataforma colaborativa de periodismo de datos que comenzó en Chile y que recientemente llegó a Colombia y Venezuela, proporciona acceso a datos públicos que contienen información sobre los vínculos entre el mundo empresarial y la política en estos países. Están surgiendo plataformas colaborativas similares y veremos más en los próximos años.

El futuro de los datos abiertos en América Latina y los desafíos futuros

Existen ventajas y algunas iniciativas que ya han comenzado a demostrar que los datos abiertos pueden contribuir al desarrollo y el bienestar de la región. Sin embargo, aunque todavía estamos en las primeras etapas, nos encontramos con tres desafíos principales por delante. El primero, desde el punto de vista de la democracia, los recientes escándalos de corrupción en Brasil, Chile, México y Argentina

10. <http://atuservicio.uy>

11. <http://www.lanacion.com.ar/data>

12. <http://excesosincastigo.convoca.pe/>

nos muestran que todavía tenemos un largo camino hacia la transparencia. Esto, sumado al aumento de la desigualdad y los problemas en la provisión de servicios públicos básicos, puede dejar atrás la agenda de los datos abiertos. En segundo lugar, a pesar de las oportunidades económicas de un gran mercado en crecimiento, la región carece de visión para desarrollar nuevos modelos de negocio basados en datos abiertos (y negocios con datos en general). Necesitamos más esfuerzos para vincular a los empresarios con el movimiento de datos abiertos y el movimiento de innovación social. Construir puentes en esas comunidades podría ayudar a liberar el valor económico de los datos abiertos. En tercer lugar, como ocurre con cualquier tecnología, a pesar de su gran potencial para promover el bien, los datos abiertos también pueden generar algún mal. Divulgar la identidad de los ciudadanos podría causar daño, discriminación o trato injusto, especialmente para las minorías o la población vulnerable. En Brasil, por ejemplo, las obligaciones impuestas por las leyes de publicación de datos personales sobre temas delicados (mujeres embarazadas y pacientes con VIH) chocan con el papel del Estado de proteger a sus ciudadanos contra la discriminación. Necesitamos una visión crítica sobre el equilibrio entre los derechos de los datos personales y el valor que puedan tener los datos abiertos publicados al detalle.

Sin lugar a duda, aunque estos son desafíos difíciles, esperamos ver más acontecimientos en el terreno de los datos abiertos en América Latina en un futuro próximo, y que estos contribuyan a generar una región más transparente, inclusiva y desarrollada.

5

Nuevas oportunidades de negocio en la Economía de los Datos

5.1

La explotación de nuevos modelos de negocio y su encaje en la regulación actual

99

5.2

La intermediación: industrias *data-driven*

106

Hoy es posible deducir los hábitos y las conductas de un individuo a partir de los datos de geolocalización que envía su teléfono móvil a las redes, y la información publicada en redes sociales. El tratamiento agregado y anonimizado de esta información permite desarrollar herramientas de enorme utilidad para la sociedad en su conjunto, y son numerosos los nuevos negocios que han nacido precisamente gracias al acceso a este tipo de información al alcance de todos, como la localización basada en redes sociales (LBRS).

Así, la potencia de las nuevas técnicas de análisis científico de datos se ha revelado como un instrumento capaz de influir en el comportamiento de las personas, de utilidad para la mejora de la planificación e incluso para la prevención de enfermedades y, con ello, infinitas oportunidades de negocio sustentadas en los datos y su explotación.

5.1

La explotación de nuevos modelos de negocio y su encaje en la regulación actual

La Economía de los Datos está permitiendo el surgimiento de nuevos modelos de negocio —tanto locales como globales— que están reconfigurando la estructura de muchos mercados y sectores productivos, los cuales pueden aumentar la eficiencia en la producción y, sobre todo, en la distribución de bienes y servicios, ya sean estos tradicionales o nuevos.

La innovación y la transparencia están desdibujando las barreras de entrada en ciertos mercados que se basaban en la información, fomentando una reorganización de los negocios tradicionales al aparecer nuevas empresas que impulsan una mayor competencia en el mercado. Buenos ejemplos de esta transformación sectorial pueden encontrarse en el sector audiovisual, en el de los medios de comunicación, en los servicios financieros, en el comercio de bienes, en el transporte y en los servicios de alojamiento.

Un elemento básico y diferencial en el que se apoyan estas nuevas empresas de la Economía de los Datos es la transparencia. Puesto que los mercados funcionan como transmisores de información —a través de los precios y cantidades de los bienes y servicios que ofrecen, y las decisiones que toman los consumidores y oferentes—, su mayor afluencia y transparencia reducen tanto los costes de transacción como los problemas derivados de la asimetría de la información. De esta forma, se introduce una mayor disciplina entre los operadores tradicionales. Un buen ejemplo de ello son los sistemas de valoración por parte de los usuarios de la experiencia de los servicios recibidos y/o de la calidad de los bienes comprados.

Estos sistemas corrigen o alinean a través de la reputación los incentivos de la oferta y la demanda, y permiten una asignación más eficiente de los recursos por parte de ambas partes del mercado. Además, a fin de mantener en la plataforma

ambos lados del mercado, los agentes deben tener confianza para interactuar o efectuar transacciones entre ellos. Por eso, las plataformas emplean por regla general distintas técnicas para atenuar las asimetrías de información, reducir los riesgos y fomentar la confianza entre los participantes, por la vía de la transparencia. Así, estipulan reglas sobre el acceso, la participación y las sanciones para los agentes que actúen a través de las mismas.

Muy directamente vinculadas a estas innovaciones se encuentran las plataformas digitales, que desde un punto de vista económico actúan como intermediarios entre oferentes y demandantes de un mercado que, al igual que en el escenario tradicional, sirven de punto de encuentro de ambas partes para la efectiva realización de transacciones e intercambios (*marketplaces*). En ausencia de estas plataformas digitales, los elevados costes de búsqueda y coordinación entre las dos partes del mercado —compradores y vendedores— impedirían que se produjesen muchas transacciones hoy posibles. La diferencia con el escenario tradicional es que las plataformas no son físicas, sino digitales, y pueden beneficiarse de una mayor flexibilidad y tamaño, a la vez que abarcar un mercado geográfico inimaginable antes de su aparición.

Las plataformas digitales ya han transformado un amplio abanico de industrias (las pioneras son aquellas relacionadas con contenidos educativos, informativos y culturales, transporte, alojamiento o compra-venta de bienes y servicios minoristas), y seguirán haciéndolo en otros sectores. La Comisión Europea reconoce una amplia diversidad de tipologías de plataformas digitales que cubren una gama de actividades que incluyen plataformas de publicidad online, *marketplaces*, motores de búsqueda, redes sociales, contenidos creativos de comunicación, plataformas de distribución de aplicaciones, servicios de comunicaciones, sistemas de pago y plataformas para la economía colaborativa.

Las plataformas aspiran a sustituir a la web como canal principal de contenidos, y sus principales características son las siguientes:

- Agregadores de contenidos y servicios: reúnen a diferentes proveedores, aplicaciones, servicios, contenidos y desarrolladores, incrementando su valor en la medida que aumenta la oferta, tanto en cantidad como en calidad, por el efecto red.
- Abiertas o cerradas: por ejemplo, Apple es una plataforma cerrada o propietaria, mientras Android o varios operadores de telecomunicaciones aspiran a crear plataformas abiertas, con diferentes grados de control sobre los desarrollos de terceros.
- Convivencia de modelos de negocio: pago por acceso, cobro por contenidos, micropagos, licencias por contenidos asociados a equipos y soportes, etc.
- Acceso diferenciado (*free*, *freemium*, *premium*) a contenidos con distintos grados de calidad de contenidos, ancho de banda, resolución, interfaz, etc.
- Diversidad de agentes en cada lado del mercado, con plataformas *business-to-business* (B2B), *business-to-consumers* (B2C) o *peer-to-peer* (P2P), fundamentalmente. Las plataformas P2P permiten reducir la infrautilización de bienes o activos ociosos y aliviar problemas sociales y medioambientales,

como la contaminación, la exclusión y la pobreza, gracias a un consumo más sostenible.

- MultisopORTE y multicanal, que permite el acceso a través de web, móviles, videoconsolas, televisiones y otros dispositivos conectados a Internet.
- Capacidad de control de contenidos, derechos de propiedad intelectual y uso a los proveedores.
- Recolectoras de datos de los clientes y gestoras del usuario final.

Por regla general, las plataformas digitales se benefician de economías de escala, ya que sus actividades de intermediación implican elevados costes fijos. Además, una característica natural de las plataformas es que cuanto mayor es el número de empresas que ofrecen sus bienes o servicios, mayor posibilidad de elección tendrán los consumidores. Es lo que se conoce como economías de red, cuya existencia no deja de constituir un importante desafío para la regulación, en cuanto habilita a las plataformas a erigir nuevas barreras de entrada a medida que las redes se multiplican bajo la misma plataforma. Esto significa que la utilidad de los consumidores en un lado del mercado aumenta conforme lo hace el número de consumidores en el otro lado. Este efecto plantea implicaciones importantes, tanto para las estrategias de negocio como para las estructuras de los mercados con múltiples lados.

Los elementos básicos que concurren en las plataformas digitales son los siguientes: (i) servicios en la nube (*cloud*); (ii) *apificación*; (iii) *open source* como acelerador; (iv) movilidad; (v) IoT, e (vi) independencia y portabilidad del *software*.

Fig. 5.1

De productos a plataformas

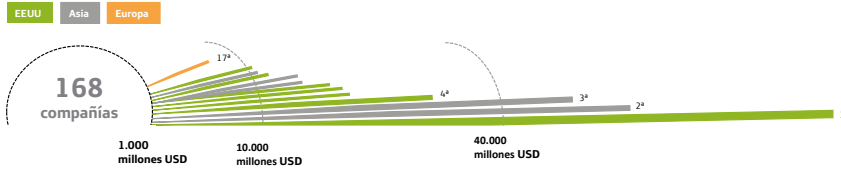
Era industrial	Era digital
<ul style="list-style-type: none"> • Productos • Cadenas de valor • Poder de control de la cadena de proveedores • Economías de escala del lado de la oferta • Activos físicos y depreciación del capital • Rendimientos decrecientes • Valoraciones de mercado vía ROA • Crecimiento orgánico o vía M&A • PIB como indicador 	<ul style="list-style-type: none"> • Plataformas • Ecosistemas • Poder de optimización de ecosistemas • Economías de escala del lado de la demanda • Activos digitales e innovación • Efectos de red • Valoraciones de mercado vía ecosistemas • Crecimiento vía efectos de red y asimetrías • Nuevas medidas, densidad digital y gratuidad

Fuente: Afi a partir de datos de Accenture.

Algunas de las plataformas digitales a nivel mundial tuvieron su origen en Europa (Skyscanner, BlaBlaCar, Spotify, etc.), pero, en su conjunto, la Unión Europea solo representa actualmente un 4% de la capitalización bursátil total de las mayores plataformas en línea, teniendo la inmensa mayoría de ellas su origen en Estados Unidos y en Asia.

Fig. 5.2

Compañías valoradas > 1.000 millones de dólares por firmas de capital riesgo



Compañía	Última valoración (miles de millones USD)	Inversión equity (miles de millones USD)	Última valoración
Uber	68,0	12,9	jun-16
Didi Chuxing	50,0	15,1	abr-17
Xiaomi	46,0	1,4	dic-14
Airbnb	31,0	3,3	mar-17
SpaceX	21,0	1,5	jul-17
WeWork	20,2	4,4	jul-17
Palantir	20,0	1,9	oct-15
Lufax	18,5	1,7	dic-15
Meituan-Dianping	18,3	4,4	ene-16
Pinterest	12,3	1,5	jun-17
Filpkart	11,6	5,9	ago-17
Lyft	11,0	3,6	oct-17
Dropbox	10,0	607	ene-14
DJI	10,0	576	sept-16
Stripe	9,2	460	nov-16
Theranos	9,0	750	feb-14
Spotify	8,5	1	abr-15
Zhong An Online	8,0	0,93	jun-15
Patym	7,0	1,8	may-17
Snapdeal	6,5	1,7	feb-16

Fuente: <http://graphics.wsj.com/billion-dollar-club/>

Nota: Solo incluye compañías privadas que han recibido inversiones en los últimos cuatro años, y cuentan con al menos un inversor de *venture-capital*.

De las 168 compañías comprendidas en el *ranking* elaborado por *The Washington Post*, 102 son estadounidenses, 46 asiáticas y apenas 15 son europeas (y solo una, Spotify, se sitúa entre las veinte primeras por volumen de capitalización).

Para aprovechar al máximo las ventajas inherentes y derivadas de la eclosión de las plataformas digitales, son varios los retos que deben abordarse, especialmente

el de la adaptación del marco regulatorio de muchos sectores para dar cabida a estas nuevas iniciativas disruptivas,¹ y garantizar unas reglas de juego básicas para todos:

- *Level playing field*, o terreno de juego equitativo, para evitar que algunas plataformas aprovechen la existencia de vacíos regulatorios (arbitraje regulatorio) para ofrecer servicios sujetos a menores requisitos que en los mercados tradicionales. Es, por tanto, preciso desarrollar el marco regulatorio que garantice tratamiento legal similar a actividades similares.
- Protección al consumidor y vigilancia de conductas responsables para proteger valores fundamentales mediante una atribución de responsabilidades clara y transparente en los procesos de intermediación que ejercen las plataformas, independientemente del modelo de negocio elegido, que determina el alcance de la interacción directa entre agentes.
- Minimización de externalidades negativas, como la gentrificación asociada a las plataformas de intercambio de viviendas, generando efectos negativos a agentes que no participan, que deben ser anticipadas y mitigadas mediante regulación y políticas públicas.
- Fomento de la competencia y de la apertura de los mercados, contrarrestando las limitaciones a la misma que puedan generar barreras de entrada y/o costes de cambio entre proveedores por la gestión de los datos. Permitir que los consumidores transfieran de forma segura sus datos entre distintas empresas, plataformas, proveedores —como ocurre a partir de enero de 2018 por efecto de la Nueva Directiva de Servicios de Pago (PSD2)— puede fomentar la competencia.
- Análisis de la competencia, en la medida en que los métodos convencionales para definir el mercado relevante o medir el poder de mercado no son aplicables en general para los mercados con múltiples lados, por los efectos de red y la ruptura de la lógica de fijación de precios atendiendo al coste marginal.
- Fiscalidad y regulación laboral, por el papel de intermediación entre las partes del mercado, incluidos los nuevos modelos de trabajo bajo demanda (*gig economy*) que demandan una revisión de las legislaciones laborales y de protección social para abarcar este nuevo modelo de relaciones.

En la revisión de medio término de la Estrategia del Mercado Único Digital, la Comisión Europea se comprometió a lo siguiente:

- Preparar acciones para abordar cuestiones relacionadas con cláusulas contractuales desleales, prácticas comerciales abusivas o resolución de controversias.
- Mejorar la coordinación del diálogo con las plataformas para establecer mecanismos y soluciones técnicas que eliminen contenidos ilícitos.
- Analizar la transparencia y responsabilidad de los algoritmos subyacentes a las plataformas digitales, en particular, sobre cómo configuran, filtran o personalizan los flujos de información que intermedian las plataformas.

1. CNMC (2016): Estudio sobre los nuevos modelos de prestación de servicios y la economía colaborativa. Madrid (CNMC 2016).

La economía de las API

Las interfaces de programación de aplicaciones (API) permiten el acceso a datos de las empresas o Administraciones públicas por aplicaciones de terceros. Una API es el mecanismo más útil para conectar dos *softwares* entre sí para el intercambio de mensajes o datos en formato estándar, convirtiéndose en un instrumento para buscar ingresos, innovar y automatizar procesos.

Hay empresas cuyo negocio proviene en gran medida de sus API. Otras empresas pertenecientes a otros sectores tradicionales (banca) están adoptando este enfoque y modificando sus procesos tanto por cumplimiento normativo (PSD2) como las nuevas condiciones de mercado, pero la gran mayoría de empresas aún no ha abordado la *apificación* de su negocio, para lo cual deben considerar lo siguiente como primer paso en la toma de decisiones:

- Entendimiento de la economía API y de los modelos de negocio *open business*, comenzando con la evaluación de los activos internos de la compañía y de la cadena de valor desde una doble perspectiva: como consumidor y como generador de valor a través API, esto es, (i) como generador de valor con el propósito de compartirlo, así como (ii) la capacidad de embeber valor e innovación de terceros dentro del propio modelo de negocio a través del canal digital API.
- Estudio de mercado y definición de la estrategia de *apificación* del negocio.
- Ubicar negocio y tecnología en un mismo plano, ya que ambos departamentos deben trabajar como un único equipo.
- Diseño y desarrollo de la arquitectura de servicios flexible y escalable que soporte la incorporación ágil y sencilla de nuevas API.
- Facilitar herramientas y cuidar a la comunidad de desarrolladores. Es cada vez más habitual que las empresas cuenten con un *API market* desde el que comparten activos con sus clientes, proveedores o socios. Esta plataforma debe ser sencilla y amigable con el objetivo fundamental de facilitar el trabajo de los desarrolladores y de otras audiencias de negocio.

La comunidad de desarrolladores es el nuevo canal de distribución, por lo que debe facilitarse lo máximo su trabajo, mediante la provisión de documentación de las API, herramientas para su integración, generación de comunidad y *engagement* para fomentar el intercambio de experiencias y la creación de valor añadido alrededor de la API.

- Seguridad y capa de gestión, fundamental para monitorizar las API disponibles, facilitar y vigilar el acceso a las mismas, controlar la facturación, etc.

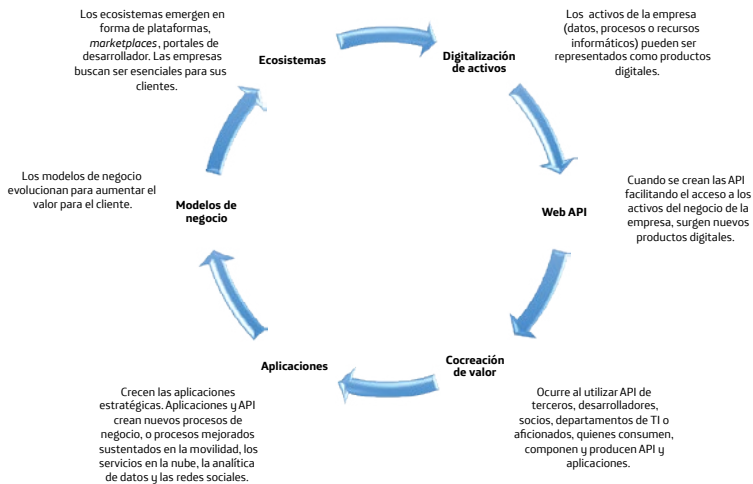
Son varias las clases de API existentes hoy:

- API de servicios web, que permiten el intercambio de información entre un servicio web (*software* que da acceso a un servicio concreto a través de una URL) y una aplicación. Normalmente ese intercambio se produce a través de peticiones HTTP o HTTPS (la versión cifrada del protocolo HTTP).

- API basadas en librerías, que permiten que una aplicación importe una librería de otro *software* para hacer el intercambio de información. Gran parte de las librerías que dan acceso a productos y servicios están diseñadas en JavaScript, como las que se utilizan dentro del mercado de la cartografía web (servicios como Google Maps, Leaflet, ArcGIS, CartoDB, MapBox o D3.js).
- API basadas en clases, que permite la conexión con los datos en torno a las clases, algo habitual en programación orientada a objetos con Java.
- API de funciones en sistemas operativos, forma habitual en que los programas de *software* interactúan con los sistemas operativos.

Fig. 5.3

Ciclo de negocio de las API



Fuente: Afi a partir de datos de IBM (2014), The Power of the API Economy Stimulate Innovation, Increase Productivity, Develop New Channels, and Reach New Markets.

Fig. 5.4

Modelos de negocio de las API



Fuente: Afi.

5.2

La intermediación: industrias *data-driven*

Una empresa basada en datos (DDC, por sus siglas en inglés) es una empresa basada en la información, donde los datos son el epicentro de los procesos y de la toma de decisiones.

Las características de las DDC o de la «cultura del dato» empresarial son las siguientes:

- Asentadas en una política de analítica de datos.
- Intensivas en uso de herramientas innovadoras de análisis de datos y de seguimiento y control de los indicadores clave de desempeño (KPI por sus siglas en inglés) de la empresa.
- Desaparición de las jerarquías tradicionales, y fortalecimiento de la comunicación directa entre dirección y empleados para una ágil toma de decisiones.
- Evaluación por objetivos.

Si bien la Economía de los Datos es susceptible de ser aprovechada por cualquier sector de la economía, los descritos a continuación han sido pioneros en adoptar —ya sea por cumplimiento normativo, por impulso innovador o por competencia de mercado— un enfoque *data-driven* en sus modelos de negocio.

Sector financiero y de seguros

El sector financiero y de seguros es un claro ejemplo de aplicación intensiva y de adopción de los principios básicos de la Economía de los Datos. La tecnología está impulsando rápidamente la transformación del sector de los servicios financieros, con enorme potencial de aumentar la competencia, la innovación y la eficiencia, y la regulación está acompañando —e incluso acelerando— este proceso. Las *fintechs* han irrumpido con fuerza en los últimos años en el sector financiero y su consolidación es cada vez más evidente.

El término *fintech* se refiere tanto a nuevas aplicaciones, como procesos o líneas de negocio que tienen el objetivo de ofrecer mejores soluciones a los servicios financieros. El ecosistema *fintech* irrumpe a lo largo de toda la cadena de valor de los servicios financieros, ofreciendo nuevos servicios financieros, añadiendo valor sobre actuales, desintermediando productos y/o servicios financieros (por ejemplo, mediante las plataformas de *crowdfunding*), produciendo y distribuyendo de forma más eficiente servicios financieros actuales y/o permitiendo mayor acceso o mayor transparencia en la oferta (por ejemplo, con los buscadores y comparadores de productos y servicios).

Fig. 5.5

Infografía del ecosistema *fintech* en España (enero de 2017)

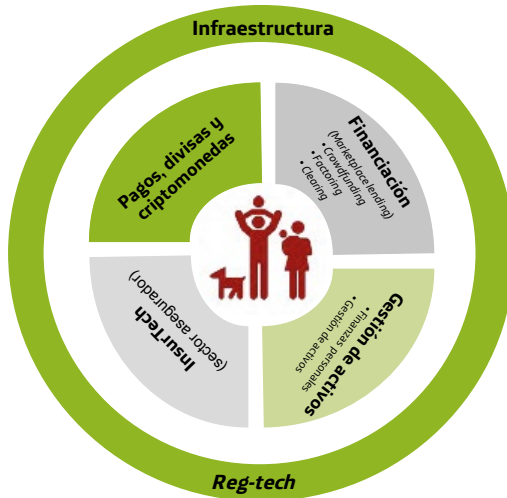


Fuente: <http://spanishfintech.net/mapa-fintech-espana/>

Aunque no existen límites definidos para la multitud de servicios y productos existentes en torno al ecosistema *fintech*, sí se pueden establecer cuatro grandes categorías en función de la finalidad principal de estos servicios.

Fig. 5.6

Categorías / Verticales del ecosistema *fintech*



Fuente: Afi.

- **Infraestructura:** medios y sistemas que facilitan o regulan los flujos de información a través de canales digitales. Incluye mecanismos que facilitan la identidad, seguridad y privacidad de las operaciones digitales, así como habilitan las transacciones con impacto monetario. La tecnología *Blockchain*, o contabilidad pública de transacciones de manera distribuida, es uno de los elementos que mayor interés ha despertado en el ámbito del sistema financiero, por su potencial aplicación a distintos ámbitos, tanto para la compensación y liquidación de operaciones como para explorar vías de diseño de una identidad digital (véase el recuadro 5.1).
- **Reg-tech:** startups que ofrecen servicios, herramientas y mecanismos digitales sustentados en datos dirigidos al conjunto del sistema financiero para favorecer y simplificar el cumplimiento normativo y de *reporting* de entidades, compañías e inversores.
- **Pagos, divisas y criptomonedas:** una mayoría de agentes del ecosistema *fintech*, ya sea desde una lógica competitiva o cooperativa con los proveedores tradicionales de servicios de pago (entidades financieras), opera en el ámbito de las soluciones de pago digitales; dentro de esta categorización, mención

aparte requiere la relevancia de las criptomonedas, cuyo máximo exponente es actualmente el bitc in.

- *Marketplaces* o plataformas de financiaci n participativa: nuevos mercados que permiten conectar prestamistas o inversores con prestatarios directamente y sin intermediaci n de una tercera parte (*trust* financiero). Se distinguen dos modelos: el predominante en el mercado actual seg n el cual la plataforma es un intermediario que simplemente pone en contacto a prestamistas y prestatarios a cambio de alg n tipo de comisi n (no financiera), y otro, menos habitual, en el que la propia plataforma tiene capacidad para hacer aportaciones en forma de deuda o capital a los prestamistas de manera directa.

Colaboraci n 4. Los datos en la reflexi n de pol tica p blica

 lvaro Mart n, economista jefe de Regulaci n Digital en BBVA Research

El proceso de transformaci n digital es una realidad observable en la mayor parte de los sectores de actividad econ mica. No se trata de un *shock* inmediato, pero s  de un proceso que conlleva profundos cambios en los modelos de negocio, las estructuras de mercado y, de manera esencial, en la relaci n de las empresas con sus clientes. El cambio suele ocurrir antes y ser m s r pido en aquellos sectores con un mayor peso relativo de los activos inmateriales sobre los bienes f sicos y tambi n, como cabr a esperar, en aquellos con menores barreras de entrada, especialmente las de tipo regulatorio.

Al tiempo que la digitalizaci n avanza imparable en todos los  mbitos de la econom a y la sociedad, se produce un fen meno de datificaci n de nuestra realidad. Nunca antes tuvimos tal cantidad de datos sobre todo lo que ocurre a nuestro alrededor, ya sea por la huella de nuestra actividad digital, por la proliferaci n de sensores conectados (Internet de las Cosas), o por la digitalizaci n y la apertura de fuentes de informaci n ya existentes. Y, al tiempo que crece la disponibilidad de datos, aumenta nuestra habilidad para extraer valor de esa informaci n, gracias a la mayor capacidad de almacenamiento, la creciente potencia de computaci n y las mejoras en algoritmia (incluyendo el campo de la inteligencia artificial).

A menudo se suele achacar el cambio disruptivo al que asistimos al auge de diversas tecnolog as que comparten su capacidad de reducir precios y alcanzar niveles de adopci n siguiendo curvas exponenciales. En este grupo selecto tendr amos la telefon a m vil, el acceso a Internet o la computaci n en la nube, por mencionar las que resultan m s familiares, por ubicuas.

Sin embargo, el acceso a estas tecnolog as es una condici n necesaria, si bien dif cilmente puede ser suficiente, para el cambio. La verdadera transformaci n est  siendo impulsada por la demanda de los ciudadanos, cuyo est ndar de calidad viene fijado hoy por la mejor oferta disponible en cada momento. Cuando todo est  al alcance de la mano a trav s de un tel fono m vil,  c mo justificar un servicio con calidad por debajo de la referencia del mercado?

En este contexto, la presión sobre los proveedores tradicionales se ve incrementada todavía más por la irrupción de proveedores que no solo ofrecen una experiencia fabulosa, sino que a menudo tienen un precio de entrada gratuito. Estos modelos generalmente adquieren la forma de plataformas digitales, esto es, son gestores de mercados multilaterales en los que se producen importantes economías de red y de alcance. En uno de los lados de las plataformas están quienes se muestran dispuestos a pagar por acceder al mercado y, en otro, los consumidores a los que los primeros quieren llegar.

La confianza como activo

El cambio de los modelos en los que el cliente paga por la adquisición o el disfrute de un bien o servicio hacia aquellos en los que son los datos generados por el cliente los que generan valor para alguna empresa, solo puede sostenerse a largo plazo si existe una relación de confianza. En los primeros estadios de transformación digital, los ciudadanos rara vez son conscientes de las dinámicas y los incentivos generados en este nuevo modelo. Es más, a menudo ignoran la abundancia y la granularidad de los datos en manos de las empresas y aún desconocen más el grado de conocimiento que pueden tener de ellos a partir de dicha información. Sin embargo, las mejoras en la regulación y la progresiva concienciación social (e individual) ante los riesgos derivados de un acceso indiscriminado a los datos personales deberían llevar a un punto de equilibrio en el que la confianza depositada en determinados actores sea la que realmente explique el nivel de información personal al que tienen acceso.

La confianza, en realidad, tiene diversas dimensiones, que pesarán más o menos dependiendo del individuo y del servicio que se espera recibir. Así, podemos confiar en la seguridad de los datos (capacidad de evitar accesos fraudulentos a la información), el respeto a la privacidad (no revelar información sensible), la explotación responsable de los datos (evitar la comercialización no deseada, por ejemplo), o el asesoramiento en el mejor interés del cliente a partir de la información conocida.

Regulación para una sociedad rica en datos

La regulación desempeña un papel fundamental en la configuración de un marco equilibrado para que el proceso de datificación lleve a mayores cotas de bienestar social. La confianza de los clientes es más fácil de obtener cuando hay un respaldo de la regulación que fija las líneas de lo que es aceptable en el tratamiento de los datos, garantiza ciertos derechos a los titulares, y ofrece mecanismos para asegurar el cumplimiento de lo dispuesto en la normativa.

En Europa, encontramos una de las sociedades en las que existe una mayor conciencia social y política acerca de la necesidad de proteger y empoderar a los ciudadanos en la gestión de sus datos personales. En 2018 será de aplicación el nuevo Reglamento General de Protección de Datos (GDPR, por sus siglas en inglés), todo un hito que reforzará los derechos y creará un marco ambicioso de protección de los datos personales de los residentes en la Unión Europea.

Entre las novedades introducidas por GDPR, podríamos destacar que se incrementa la exigencia en la recolección del consentimiento y su trazabilidad en el ciclo de vida del dato; se introduce la privacidad por diseño y por defecto como principios, y se incluye un refuerzo de las sanciones aplicables en caso de incumplimiento por parte de quienes procesan dichos datos. Además, el nuevo reglamento reconoce un derecho de portabilidad que se suma a los ya reconocidos derechos de acceso, rectificación, cancelación y oposición. Se espera que la portabilidad permita a los ciudadanos evitar situaciones de acceso exclusivo por parte de algunas empresas digitales (que pueden llegar a producir efectos de *lock-in*), pero además tendrá consecuencias económicas de gran calado.

Conflictos entre regulaciones sobre datos

Los flujos digitales han acelerado un proceso de globalización que se inició con el comercio internacional, y ahora permite la prestación de servicios de manera remota y el acceso a información en tiempo real con independencia de su ubicación. En estas condiciones, la falta de coordinación internacional en una materia tan sensible como la protección de datos puede generar distorsiones indeseables. Esto es achacable tanto a la relativa novedad de la materia y la complejidad de la misma, como a la ausencia de un criterio homogéneo entre territorios respecto a la importancia y forma de proteger los datos personales.

Por otro lado, lo digital no solo es global por naturaleza, sino que además es crecientemente transversal. Las tecnologías, los conocimientos y los mercados tienden a confluir, desdibujando así las delimitaciones previas entre sectores. Sin embargo, persisten los marcos verticales de regulación (esto es, por actividad) que también afectan a la manera en que se han de tratar los datos. Sus objetivos suelen parecer complementarios con el marco general de protección de datos (velando por la protección al consumidor o la competencia, por ejemplo). Estas regulaciones no tienen en cuenta cómo las experiencias digitales integran cada vez más actividades que tradicionalmente se englobaban en sectores bien diferenciados (pensemos en todo lo que se puede hacer hoy a través de una red social, por ejemplo) y, sin embargo, tardan en ser reguladas a través de dichos marcos verticales, demostrando una falta de visión transversal por parte de las autoridades.

En el caso del sector financiero, la necesidad de generar ecosistemas abiertos en los que se incremente la competencia sin afectar a la estabilidad financiera ha llevado a buscar la compartición con terceros de la información de cuentas de pago de clientes. La nueva Directiva Europea de Servicios de Pago (PSD2, por sus siglas en inglés) ofrece, frente al marco general de portabilidad de GDPR, una manera más estandarizada, mejor resuelta técnicamente (gestión del consentimiento, seguridad de la información, comunicación directa entre empresas), y con la posibilidad de acceder casi en tiempo real a los datos de las entidades bancarias.

Aunque el marco vertical introducido por PSD2 se mantiene en la lógica de entidades reguladas del sector financiero (al fin y al cabo, los terceros deben ser entidades registradas ante los reguladores), en la práctica se reduce al mínimo la barrera de

entrada para que empresas de otros sectores puedan beneficiarse (con el consentimiento de los clientes, por supuesto) del inmenso valor de los datos transaccionales de cuentas de pago. Esto genera una enorme asimetría con otros sectores que también manejan grandes cantidades de datos observados de sus clientes.

El efecto combinado de PSD2 y GDPR da muestra de la importancia de que las autoridades adopten una visión holística de los datos, evitando solapamientos regulatorios que generen asimetrías entre sectores. En última instancia, de esto dependerá que se desarrolle una sana competencia entre proveedores digitales innovadores que aspiren a satisfacer las necesidades de los clientes que hayan depositado en ellos su confianza.

- **Gestión de activos:** las nuevas herramientas de predictibilidad de mercado y avances tecnológicos han abierto la puerta de entrada a gestores y *traders low cost (robo advisors)* que permiten la creación de una nueva oferta de producto, incrementan la accesibilidad a este tipo de asesoramiento y productos de inversión al disminuir las comisiones y se apoyan en la gestión de datos (*big data*) para implementar su estrategia de inversión.
- **Insurtech:** engloba el efecto de la aplicación de tecnología para mejorar la eficiencia del sector asegurador e incluso permitir la comercialización de nuevos productos.



Recuadro 5.1 *Blockchain*: paradigma de descentralización de la Economía de los Datos

Blockchain es una tecnología de contabilidad distribuida (DLT, por sus siglas en inglés) que llegó en torno a 2009 materializado en un activo digital considerado por muchos como moneda virtual o criptomoneda —el bitcóin (BTC) — que asumió el protagonismo asociado a dicha tecnología.

Si bien la criptomoneda ha sido hasta la fecha la aplicación más aterrizada y popular de esta tecnología sobre cuyo protocolo (Bitcoin, con mayúsculas) se han desarrollado miles de criptomonedas alternativas —unas compiten y otras colaboran con bitcóin—, la creación de un sistema electrónico de pago entre pares no es más que una de las múltiples aplicaciones que se están desarrollando en la actualidad, la mayoría de ellas en fase de prueba de concepto y sustentadas en contratos inteligentes o *smart contracts* (SC) y *tokenización* de activos.

En términos globales, IBM prevé que el 15% de los bancos mundiales estará utilizando *blockchain* a finales de 2017, y un 66% a finales de 2021, pero sus aplicaciones son incluso más interesantes en otras industrias, tales como la energética, sanitaria y educativa, entre muchas otras.

Del análisis realizado por William Mougayar, experto en *blockchain* y autor de *The Business Blockchain*, a finales de 2016 existían 25 consorcios globales desarrollando diversas iniciativas *blockchain*, con más de quinientos cincuenta miembros de distintas industrias, siendo la financiera la más representativa. En 2017, sin embargo, se observa una mayor presencia de consorcios multisectoriales, como el recientemente anunciado en España —Red Lyra, hoy Alastria—, primera plataforma tecnológica multisectorial del mundo, basada en *blockchain*, con todos sus nodos en territorio español para garantizar la legalidad. Alastria tiene como objetivo desarrollar de forma colaborativa un sistema de identidad digital, basado en *smart contracts*, donde, ajustándose al protocolo Ethereum, cada participante podrá desarrollar aplicaciones de manera competitiva.

Respecto a Ethereum, este protocolo no fue concebido, a diferencia de Bitcoin, para servir como una red de pagos. Su criptomoneda nativa —*ether* (ETH) — es la moneda de pago creada para actuar como «gasolina» para la máquina virtual de Ethereum y está destinada a premiar a los desarrolladores, permitir que los clientes paguen por las aplicaciones y las operaciones que se desarrollan dentro de la plataforma y evitar el *spam*. Busca promover la I+D+i y la educación para universalizar los protocolos y las herramientas descentralizadas a través de la colaboración con *desarrolladores* y conseguir universalizar las aplicaciones descentralizadas, los *smart contracts* y las organizaciones autónomas descentralizadas (DAO, por sus siglas en inglés) en torno a la tecnología *blockchain*, desarrollado originalmente como una versión mejorada del protocolo Bitcoin. Ethereum cuenta con más de setenta mil desarrolladores que de forma colaborativa —a cambio de *ether*, la gasolina de Ethereum— crean nuevos protocolos en código abierto, a disposición de la sociedad para agilizar el tránsito del Internet de la Información al Internet del Valor.

Fuente: Afi Empresa Global, n.º 174, mayo de 2017.

Sector energético y de la movilidad

El suministro y el consumo de energía confiable, asequible, eficiente y no contaminante es el reto que enfrentamos como sociedad del siglo XXI. Las aspiraciones en materia de autonomía energética y/o eficiencia energética a través de innovaciones tecnológicas tales como *smart grid* o redes inteligentes, gestión de la demanda, generación renovable y adopción de vehículos eléctricos, pueden mejorar sustancialmente la calidad de vida de las personas haciendo más asequible el acceso a la energía, reduciendo los costes de acceso y facilitando la autonomía. Innovaciones que, necesariamente, pasan por una transformación de sectores tradicionales y altamente regulados, como el energético, a un enfoque de DDC.

Desde el punto de vista de la generación de energía, la Economía de los Datos permite a las empresas del sector mejorar la eficiencia operativa, que se ha caracterizado habitualmente por ser muy costosa. Sirva de ejemplo la posibilidad de que las aspas de un molino de viento giren en función de la dirección del viento para poder generar la máxima energía posible en cada momento. Una posibilidad que ha sido fruto no solo de la sensorización del molino, sino también del procesamiento y análisis de los datos que ha generado a lo largo de su vida útil.

Desde el punto de vista del consumo de energía, la instalación de contadores inteligentes y la analítica que favorece son un claro ejemplo del avance que ha propiciado la Economía de los Datos en el sector energético. En concreto, con estos contadores se puede hacer una medición instantánea del consumo de energía y de realizar una recomendación personalizada al cliente, por ejemplo, verificando si la potencia contratada es la que realmente necesita, controlando el consumo energético de segundas residencias, o contratando planes de consumo específicos que permiten distribuir el consumo eléctrico hacia los momentos del día en los que el coste del suministro es más reducido («horas valle»).

Los nuevos modelos de negocio en el ámbito de la movilidad son numerosos —algunos de ellos inspirados en la economía colaborativa-circular—, como también lo son las herramientas y aplicaciones de geolocalización, etc. que están facilitando el inicio de una nueva era de la movilidad. Tanto es así que las personas ya no necesitan poseer un vehículo para tener uno a su disposición a su conveniencia (gracias a modelos de negocio como los del *carsharing*, *carpooling*, *bikesharing*) y los fabricantes de vehículos se están transformando gradualmente de meros fabricantes a empresas de servicios de movilidad, totalmente inmersas en la Economía de los Datos.

Las encuestas origen-destino de viajeros (O-D) han sido tradicionalmente una herramienta fundamental de generación de bases de datos con los patrones de viajes para la formulación de planes y proyectos de movilidad de personas y mercancías. Hoy las encuestas son complementadas —e incluso sustituidas, cuando los datos de las distintas plataformas son abiertos— por el análisis científico de datos (*big data*) capturados por distintos medios como las validaciones de los títulos de transporte en los puntos de acceso y salida y las señales de los dispositivos móviles de los conductores, entre otras técnicas innovadoras. La mayor conectividad actual

y la consecuente generación de datos permiten a las autoridades públicas gestionar el tráfico y las infraestructuras del transporte (redes públicas de carga de vehículos eléctricos, sistemas de aparcamiento inteligente, mapas de tráfico) en tiempo real para contribuir a una movilidad más eficiente.

El sector de la movilidad es, por tanto, un exponente de las DDC y el futuro inminente estará protagonizado por la conducción autónoma, sustentada en (i) los desarrollos de la inteligencia artificial y el *machine learning* y en (ii) el desarrollo de innovadoras capacidades de almacenamiento de energía eléctrica (baterías) para garantizar la autonomía del vehículo autónomo.

Sector comercio

En el sector del comercio minorista son muchas las aplicaciones informáticas y plataformas digitales (*marketplaces*) que han aparecido a la vez que se ha ido desarrollando la Economía Digital y Economía de los Datos y que han cubierto un hueco hasta ahora no abordado por la distribución comercial más tradicional. Desde aquellas empresas que tratan de poner en contacto directo a los fabricantes de los productos con el consumidor final, hasta aquellas que simplemente facilitan las transacciones entre ambas partes del mercado sin renunciar al eslabón tradicional de la cadena que representan los grandes distribuidores comerciales.

Independientemente de la naturaleza y la función de cada una de estas empresas, lo cierto es que la Economía de los Datos, las herramientas del *big data* y de *data science* permiten también la emergencia del importante análisis del comportamiento de los consumidores. Esta capacidad analítica del comercio minorista depende de la recopilación de los datos en la tienda, los relativos al producto y al cliente. La posibilidad de explicar, anticipar y recomendar comportamientos del consumidor en tiempo real es la verdadera revolución del empleo de la Economía de los Datos en este sector económico. El papel del *marketing* digital, en este sentido, es crucial para garantizar el éxito de estas iniciativas.

Pero, al igual que ocurre con otros sectores, el valor de esta información no se circunscribe al sector del comercio, sino que puede extenderse a otros, como, por ejemplo, el sector inmobiliario, en la medida en que la configuración, la ocupación y el valor del espacio comercial dependen del volumen de facturación que sea capaz de generar la empresa que se instale o ubique en el mismo.

Sector salud

Un área donde han sido posibles importantes avances a través del *big data* es el sector de la salud, si bien es un área en que los datos de carácter personal son particularmente sensibles (como reconoce el artículo 9 del RGPD).

El *big data* puede utilizarse para predecir la probabilidad de desarrollar una enfermedad rara; para vigilar los efectos adversos relacionados con productos farmacéuticos y dispositivos médicos; para proporcionar diagnósticos más rápidos (mediante el uso de dispositivos portátiles); para ofrecer soluciones personalizadas de salud

o para mejorar la eficacia de los tratamientos. Por ejemplo, la Iniciativa Europea de Reporte Avanzado sobre Fármacos (EU-ADR, por sus siglas en inglés) automatiza el análisis de los datos almacenados en las historias médicas electrónicas de más de treinta millones de ciudadanos europeos, con el objeto de monitorizar los efectos y la seguridad de medicamentos utilizados con poca frecuencia. El *big data* posibilita, en este caso, detectar matices en las subpoblaciones que no serían evidentes en muestras de menor tamaño.

Los algoritmos de *machine learning* también son utilizados por investigadores del sector salud para predecir, por ejemplo, paros cardíacos o respiratorios, minimizando los riesgos de este tipo de episodios traumáticos entre la población y sus consecuencias en términos de coste de los cuidados. El binomio datos-salud es fundamental especialmente en un contexto como el europeo, en que la población de más edad será mayoría en las próximas décadas, con necesidades de atención médica por condiciones de cronicidad o reducción de la movilidad, sin comprometer la calidad del servicio que se presta a los cuidados.

Administraciones públicas

El gobierno abierto u *open government* tiene como objetivo que ciudadanos, empresas, desarrolladores, tercer sector, etc., colaboren en la creación y la mejora de los servicios públicos, y en el fortalecimiento de la transparencia y la rendición de cuentas.

La Alianza para el Gobierno Abierto (AGA) / *Open Government Partnership* (OGP) nació en 2011 por voluntad de ocho Gobiernos fundadores (Brasil, Indonesia, México, Noruega, Filipinas, Sudáfrica, Reino Unido y Estados Unidos) con el objetivo de proveer una plataforma internacional a través de la cual los Gobiernos de los setenta y cinco países miembros y quince gobiernos subnacionales se comprometen a implementar cuatro principios básicos de un gobierno abierto: transparencia y acceso a la información, rendición de cuentas, participación ciudadana e innovación en los asuntos públicos a través de las nuevas tecnologías.

AGA busca que los Gobiernos sean más transparentes, rindan cuentas y mejoren la capacidad de respuesta hacia sus ciudadanos, lo que requiere un cambio de normas y cultura para garantizar un diálogo y una colaboración genuinos entre Gobierno y sociedad civil, que puede consultarse en el Manual para Puntos de Contacto de Gobierno.

6

El auge de nuevas profesiones

6.1

Nuevos liderazgos: director de datos - *chief data officer*

123

6.2

Ingeniero de datos - *big data engineer*

124

6.3

Científico de datos - *lead data scientist*

125

6.4

Director de seguridad de la información - *chief information security officer*

126

6.5

Delegado de protección de datos - *data protection officer (DPO)*

130

La revolución digital que supone el desarrollo de la Economía de los Datos ha abierto el debate del impacto que ejercerá sobre el empleo.

En la medida en que las máquinas puedan reemplazar parte o la totalidad de algunas tareas desarrolladas actualmente por trabajadores, gracias al procesamiento de información de manera instantánea (*machine learning*) o a los nuevos avances en la producción industrial (como la impresora 3D o el vehículo autónomo), resulta inevitable pensar que buena parte de los puestos de trabajo que conocemos hoy desaparecerán o, como mínimo, se transformarán.

Sin embargo, la revolución digital también abre un importante y variado abanico de oportunidades laborales que es conveniente conocer y aprovechar. En torno a las nuevas áreas de conocimiento digital y la explotación de las potencialidades que brinda este nuevo entorno, emergen nuevas profesiones digitales, cada vez más especializadas, que se clasifican según las áreas de trabajo en las que se enmarcan en la actualidad:

1. Estrategia digital:
 - a) *CDO / Digital manager*. Responsable de estrategia digital.
 - b) *Digital communication director (DCD)*. Responsable de comunicación digital.
 - c) *Digital transformation project manager (DTPM)*. Responsable de proyectos digitales.

2. Marketing digital:
 - a) *Digital marketing manager*. Responsable de *marketing* digital.
 - b) *Inbound marketing specialist*. Experto en *inbound marketing*.
 - c) *SEM y SEO specialist*. Experto en SEM y SEO.
 - d) *SEO specialist*. Experto en SEO.
 - e) *Social CRM manager*. Responsable de relación digital con clientes.
 - f) *RTB manager*. Responsable de conversión y captación programática.
 - g) *Traffic manager*. Responsable de tráfico web.

3. Contenido digital:
 - a) *Branded content specialist*. Experto en contenido digital de marca.
 - b) *Content manager*. Responsable de contenido digital.
 - c) *Copywriter*. Experto en redacción publicitaria.

4. Social media:
 - a) *Community manager*. Responsable de gestión de comunidades y redes sociales.
 - b) *Social media manager*. Responsable de la estrategia de medios sociales.

5. *Big data* y *business analytics*:
 - a) *Chief data officer*. Director de datos.
 - b) *Lead data scientist*. Científico de datos.

- c) *Big data engineer*. Ingeniero de datos.
 - d) *Visual data scientific*. Responsable de visualización de datos de negocio.
 - e) *Customer intelligence analyst*. Analista de inteligencia de clientes.
6. *eCommerce*:
- a) *eCommerce manager*. Responsable de comercio electrónico.
 - b) *eCommerce analyst specialist*. Experto en analítica para comercio electrónico.
7. Recursos humanos:
- a) *eRecruitment manager*. Responsable de reclutamiento digital.
 - b) *Human resources analyst*. Analista de recursos humanos.
8. *Digital sales*:
- a) *Digital account manager*. Ejecutivo de cuentas digitales.
9. Ciberseguridad:
- a) *Chief information security officer (CISO)*. Director de seguridad de la información.
 - b) *Data protection officer (DPO)*. Delegado de protección de datos.

La demanda de estos perfiles digitales no solo ha crecido de forma exponencial en los últimos años, sino que está llamada a ejercer un papel relevante en la nueva estructura del mercado de trabajo.

En 2016, según las estimaciones de IDC y Open Evidence para la Comisión Europea,¹ había 356.000 trabajadores en España desempeñando tareas asociadas a la Economía de los Datos en prácticamente todos los sectores económicos. En términos relativos al total de trabajadores de cada sector, su penetración se reducía apenas al 2,3%, ligeramente inferior a la media de la UE-28 (3,1%).

Las previsiones de IDC y Open Evidence para el periodo 2014-2020 apuntan a un incremento de los trabajadores españoles asociados a la Economía de los Datos hasta las 413.000 personas (se prevé, por tanto, un crecimiento medio anual del 3,8%). Un aumento que, aunque es inferior al previsto para la media europea (crecerá, en media, un 6,1% cada año), supone un avance significativamente mayor al 1% previsto por Cedefop,² organismo dependiente de la Comisión Europea, para el conjunto de los trabajadores, por lo que parece bastante evidente que, si se cumplen estas perspectivas, los trabajadores asociados a la Economía de los Datos adquirirán un mayor protagonismo en el mercado de trabajo español.

Actualmente, las profesiones digitales asociadas al *marketing* son las más demandadas en el mercado laboral,³ pero ¿cuáles son las que más van a crecer en los

1. IDC y Open Evidence (2017): European Data Market Smart 2013/0063.

2. Cedefop es el Centro Europeo para el Desarrollo de la Formación Profesional.

3. Inesdi Digital Business School (2017): Top 25 profesiones digitales 2017.

próximos años? Según el área de trabajo desempeñada, una buena muestra de directivos de grandes empresas españolas⁴ apunta a que las profesiones digitales que más aumentarán en el futuro serán las asociadas al *big data* y la ciberseguridad.

De hecho, otros estudios⁵ comienzan a constatar cierta carencia de profesionales cualificados que satisfagan las necesidades de las empresas, lo que está provocando un aumento del número de vacantes de empleo, especialmente en el caso de los desarrolladores de aplicaciones móviles, de soluciones de *big data* o los especialistas en ciberseguridad. Una carencia de profesionales que sorprende, al coincidir con la abundancia de personas desempleadas de nuestro país (España tiene la segunda tasa de paro más alta de la UE-28), y que exige el reforzamiento del sistema educativo y de las políticas activas de empleo, sobre la base de las habilidades requeridas por estos nuevos puestos de trabajo.

A continuación, se describen la misión, función y formación de los siguientes profesionales digitales:

- Director de datos, por ser el más relevante en la nueva jerarquía laboral que emana de la Economía de los Datos.
- Científico de datos e ingeniero de datos, por ser los ocupados que más están aumentando recientemente.
- Director de seguridad de la información y delegado de protección de datos, por ser unos de los que más van a crecer en los próximos años si se cumplen las previsiones.

6.1

Nuevos liderazgos: director de datos - *chief data officer*

El director de datos, más conocido por sus siglas en inglés como CDO (*chief data officer*), es el máximo responsable de los equipos especialistas en *big data* dentro de la empresa. Concretamente, es la persona que lidera la gestión y el análisis de los datos. Su misión consiste en asegurar la calidad de los datos ante los diversos reguladores y la generación de valor a los distintos niveles ejecutivos de la empresa. Entre sus funciones, se encuentran las siguientes:

- Diseñar las políticas y procesos dentro de la empresa para gestionar la calidad, consistencia, uso, seguridad y disponibilidad de los datos.
- Identificar y coordinar a los responsables de la supervisión de la ejecución de dichas políticas y procesos.
- Dibujar la arquitectura de los datos y la información para estandarizar los pro-

4. Fundación PwC y Fundación Atresmedia (2016): *Empleos del futuro en el sector audiovisual*.

5. Infojobs y ESADE (2017): *Estado del mercado laboral en España*.

cesos dentro de la empresa. Debe conocer cómo se genera, cómo se integra, cómo se utiliza y con qué tecnología se procesa la información de la empresa.

- Explorar y presentar a los responsables de las áreas de negocio de la empresa las posibilidades del análisis de la minería de datos que permita detectar comportamiento de clientes, entre otros usos de la información.

Este tipo de profesionales suele aglutinar una serie de conocimientos formativos y habilidades profesionales, entre las que destacan las siguientes:

- Conocimientos universitarios o superiores de gestión y administración de empresas (orientado a negocio), de *marketing* y de las tecnologías digitales (generalmente, licenciaturas en informática o telecomunicaciones).
- Experiencia en la creación, la implantación y la monitorización de los planes de estrategia digital en la empresa.
- Liderazgo directivo y sensibilidad hacia la innovación y el emprendimiento en la empresa.

6.2

Ingeniero de datos - *big data engineer*

El ingeniero de datos o *big data engineer* es la persona que se encarga de la identificación, la limpieza y la clasificación de los datos que serán analizados e interpretados por la figura de los científicos de datos, o *data scientists*. Su misión es preparar la información para su posterior procesamiento y puesta en valor. Entre sus funciones, se encuentran las siguientes:

- Construir el *software* necesario para integrar la información de los sistemas fuente en la plataforma *big data* de la empresa, de tal forma que quede preparada para su uso.
- Extraer e integrar los datos (estructurados o no) procedentes de diversas fuentes de información (internas o externas) en grandes conjuntos de datos.
- Consolidar, limpiar y clasificar los datos para su posterior uso en informes y análisis particularizados.
- Administrar y mantener actualizada la base de datos de la empresa.
- Optimizar el rendimiento del ecosistema de *big data* de la empresa.
- Proporcionar un acceso simplificado a los datos para los usuarios de los mismos.

Este tipo de profesionales cuenta con la siguiente formación y habilidades:

- Titulados universitarios, generalmente, en informática o telecomunicaciones.

- Manejo del lenguaje de programación con diversos paquetes informáticos, en particular, de aquellos que permiten el almacenamiento y la lectura de *big data*.
- Buena comprensión de herramientas de extracción, transformación y carga de información que servirán de base para crear y gestionar trabajos relativos a la misma.

6.3

Científico de datos - *lead data scientist*

El científico de datos o *lead data scientist* es el trabajador encargado de convertir importantes volúmenes de datos (generalmente, desestructurados) en información útil para la empresa, mediante técnicas de análisis avanzadas que permiten detectar y predecir el comportamiento de los consumidores. Su misión es analizar datos cuantitativos y cualitativos para extraer información de valor que ayude a la empresa a tomar decisiones para mejorar la experiencia de los clientes. Entre sus funciones, se encuentran las siguientes:

- Dar sentido y significado a los datos que se recogen a través de los sistemas de almacenamiento de *big data*.
- Analizar estadísticamente la información y detectar comportamientos de los consumidores que generen valor para la empresa.
- Desarrollar modelos predictivos del comportamiento de los consumidores.
- Representar gráficamente grandes volúmenes de datos, crear cuadros de mando y elaborar informes que sirvan para la toma de decisiones estratégicas de la empresa.

Entre las competencias formativas y la experiencia que se requieren en la actualidad a este tipo de profesionales, se encuentran las siguientes:

- Titulados en disciplinas científicas, relacionadas con las matemáticas, estadística, informática, telecomunicaciones, etc.
- Mezcla de conocimientos técnicos (control de las nuevas tecnologías, programación informática, manejo de paquetes estadísticos) y estadísticos (análisis descriptivo y predictivo).
- Habilidades comunicativas para trasladar resultados obtenidos de forma visual (análisis gráfico, infografías, mapas) y conocimiento del negocio.

6.4

Director de seguridad de la información - *chief information security officer*

El director de seguridad de la información es el responsable de proteger la información digital de los sistemas interconectados. El desarrollo de la Economía de los Datos implica retos en ciberseguridad que requieren la protección de la información procesada, almacenada y transportada por los sistemas, y el tratamiento de las amenazas. Su misión consiste en diseñar y ejecutar el plan de ciberseguridad de la empresa que garantice la disponibilidad, la integridad y la confidencialidad de los datos. Entre sus funciones, se encuentran las siguientes:

- Formular planes para garantizar la disponibilidad de la información, la propiedad intelectual, el cumplimiento normativo y los archivos informáticos de la empresa.
- Clasificar y categorizar las amenazas y vulnerabilidades de la información.
- Diseñar un plan de contingencias para preparar a la empresa ante las ciberamenazas que pueda sufrir.
- Identificar los ciberataques y responder a los mismos.

Los conocimientos y habilidades que deben reunir estos profesionales digitales son, además de las directivas, los siguientes:

- Estudios superiores (al menos, universitarios) en informática o telecomunicaciones, además de haber completado cursos de especialización en el ámbito de la seguridad digital y la auditoría.
- Experiencia en la gestión de riesgos, en áreas de seguridad corporativa y auditoría.
- Conocimiento de las mejores prácticas de seguridad de la información.

Colaboración 5. El auge de las nuevas profesiones

Alberto Turégano, asesor de arquitecturas *business intelligence & big data*

Los últimos años han traído cambios que están transformando las vidas de las personas y las capacidades de los objetos. Vivimos en una sociedad en la que todo se está «hiperconectando» y la creciente digitalización, tanto de las personas como de las cosas (*Internet of everything*), está propiciando que el 90% de los datos almacenados hoy en todo el mundo se haya generado únicamente durante los dos últimos años. Diversos estudios señalan que para 2020 se habrá multiplicado por diez el volumen de datos que se genera anualmente en el planeta, y el tratamiento inteligente del mismo será clave para mejorar la eficiencia de los diferentes servicios de tal forma que esta redunde, en definitiva, en una mayor aportación de valor a las personas.

¿Qué acciones hay, por tanto, que poner en marcha para que empresas e instituciones obtengan el máximo valor del ingente volumen de datos que se está generando? Lo primero de todo, y parece obvio, es almacenarlo. Y, tras almacenarlo, es necesario el poder procesarlo con cierta eficiencia, de tal forma que se deje listo para ser analizado. Para ello es preciso que las organizaciones se adapten a este nuevo entorno digital y se transformen internamente, lo que conlleva abordar los siguientes puntos:

- Inversión en tecnología. La era del *big data* proporciona la oportunidad de afrontar con ciertas garantías el reto de un nuevo entorno, con mayor volumen de datos, mayor variedad en su tipología, así como una mayor velocidad tanto en la generación como en la necesidad de procesamiento de los mismos.
- Necesidad de una transformación cultural en los equipos. La resistencia al cambio suele ser el gran talón de Aquiles de gran número de organizaciones. Fomentar la adopción de nuevas capacidades digitales para los integrantes de la organización se presenta como el elemento clave para conseguir el éxito en la transformación digital.
- Adopción de nuevos perfiles. Para poder abordar con garantías la cadena de valor de la información comentada anteriormente, resulta imprescindible contar con nuevos perfiles dentro de la organización, perfiles que se verán con más detalle a continuación.

Arquitecto de *big data*

El arquitecto de *big data* es el responsable de diseñar e implementar la arquitectura necesaria para dar respuesta al caso de uso concreto propuesto por el cliente. Dado que los requerimientos de los diferentes casos de uso pueden diferir en gran medida unos de otros tanto en las necesidades de velocidad de los datos (unos pueden necesitar obligatoriamente la integración de los mismos en tiempo real mientras que para otros casos de uso esta latencia puede ser optativa) como en la tipología de los mismos (en unos casos se tratarán datos no estructurados como *tuits*, semiestructurados como *logs* o totalmente estructurados como transacciones en un sistema operacional), el arquitecto de *big data* deberá tener un profundo conocimiento de las diferentes tecnologías y la relación entre ellas, con el objetivo de integrarlas y combinarlas adecuadamente y poder así solventar el problema de negocio presentado. También será responsable de implementar las políticas de seguridad y gobierno necesarias para cumplir con las diferentes normativas.

Ingeniero de datos

Si bien los arquitectos de datos son los encargados de implementar la arquitectura adecuada, es decir, los encargados de seleccionar la tecnología, los ingenieros de datos son los responsables de construir el *software* necesario para integrar la información de los sistemas fuente en la plataforma *big data*, de manera que quede preparada para su uso por parte de los científicos de datos de forma correcta y con



la máxima calidad posible para su análisis. Este perfil surge como necesidad a una situación que se estaba dando en el pasado (aunque se sigue dando con cierta frecuencia en la actualidad en algunas organizaciones), por la cual los científicos de datos dedicaban un 80% de su tiempo al proceso de limpieza y almacenamiento de datos (proceso conocido como *data wrangling*), perdiendo el foco en lo que debería ser su actividad principal: la analítica avanzada.

Por tanto, los ingenieros de datos deben ser conocedores de los principales lenguajes de programación utilizados en los distintos casos de uso, así como dominar las disciplinas de calidad e integración del dato.

Científico de datos

Este es posiblemente el perfil más complicado de encontrar en el mercado, o de desarrollar internamente dado el caso. El científico de datos es mejor en estadística que cualquier ingeniero de *software*, y mejor ingeniero de *software* que cualquier estadístico, como apuntó Josh Wills en 2012. Su principal responsabilidad es la de extraer conocimiento de la información proporcionada por los ingenieros de datos para poder así responder a las preguntas que se le plantean en los diferentes casos de uso.

El científico de datos, «el trabajo más sexi del siglo *xxi*» como ya se le conoce a este papel, debe poseer, además de conocimientos de ciencia aplicada, las aptitudes

necesarias de comunicación y de negocio de tal forma que sea capaz de trasladar con claridad a los responsables de la toma de decisiones los *insights* obtenidos durante el proceso de análisis. Una parte muy importante de la comunicación tiene que ver con las capacidades de *storytelling* y visualización de datos, por lo que se suele exigir también conocimientos en herramientas de visualización.

En la realidad, es tan difícil encontrar estos perfiles en el mercado que en el sector se los conoce como unicornios. No obstante, la demanda de estos perfiles por parte de la industria no ha parado de crecer desde 2011, y en el portal de búsqueda de empleo Glassdoor, se colocó al científico de datos en el número uno de la lista «los 25 empleos más prometedores de 2016».

CDO (*chief data officer*)

Debido a la explosión de datos comentada con anterioridad, la mayoría de las organizaciones se encuentran con el problema de gobernar esta información. En algunos casos, dichas organizaciones se han preocupado de realimentar los sistemas operacionales con los *insights* obtenidos del *data warehouse* en relación con la calidad del dato, y actualmente dicha información satisface los estándares de calidad, pero se han despreocupado de la manera en que se manipulan, transforman o agregan posteriormente estos datos en los entornos analíticos. Todo ello puede llevar a inconsistencias que redundarían en errores en la elección de la estrategia correcta por parte de los responsables: errores que llegarían a ocasionar graves consecuencias para el negocio.

Por ello, surge el papel de CDO, ejecutivo responsable de gobernar todos los datos de la empresa, y de definir las diferentes políticas de gobierno en relación con la estrategia, control y explotación eficiente de la información con el objetivo de crear valor para negocio.

En una encuesta realizada por el IBM Institute for Business Value, un tercio de las organizaciones encuestadas contaban con este puesto en verano de 2015 y, según Gartner, se estima que el 50% de las empresas en las industrias reguladas tendrán un CDO para 2018 y este número alcanzará el 90% para 2019, lo que supone sin duda alguna una oportunidad de desarrollo profesional en años venideros.

Conclusión

El hecho digital lo está transformando todo y, aunque cada vez estamos más familiarizados con las nuevas tecnologías, se ha abierto todo un abanico de oportunidades para la obtención de un empleo relacionado con el análisis masivo de datos en grandes corporaciones. Tareas que tradicionalmente realizaban personas con formación en ingeniería informática, matemáticas o físicas, se pueden llevar a cabo también hoy por alumnos de posgrados que imparten diferentes escuelas de negocios o empresas de formación orientadas a este ámbito. Según la consultora McKinsey, habrá una demanda de millón y medio de empleos relacionados con los puestos descritos anteriormente durante el próximo lustro. ¡Bienvenidos a la profesión más atractiva del siglo XXI!

6.5

Delegado de protección de datos - *data protection officer* (DPO)

Es la nueva figura requerida por el RGPD para el adecuado ejercicio de las funciones detalladas en la norma, y como punto de contacto con interesados, responsables y autoridades de control.

Si bien algunas entidades ya contaban con esta figura, en muchos casos sus funciones y/o configuración organizativa no se corresponden con lo exigido por el nuevo reglamento, pues estaban originalmente ubicados en las áreas legales o de seguridad de la información, cuando parte de las funciones que absorberá el DPO se desempeñan en otras áreas de las entidades. La AEPD está trabajando con la Entidad Nacional de Acreditaciones en la certificación de DPO como garantía de cualificación y capacidad.

El perfil profesional del DPO deberá responder a cualidades profesionales y a sus conocimientos especializados de la legislación y las prácticas en materia de protección de datos, así como a su capacidad para ejecutar las tareas contempladas en el artículo 37 del Reglamento. Aunque solo se mencionan conocimientos en legislación sobre protección de datos, el DPO deberá tener conocimientos sobre seguridad informática, esquema nacional de seguridad y seguridad de la información.

Para salvaguardar su independencia, el Reglamento establece una serie de obligaciones para los responsables y encargados de tratamiento: dotación de medios materiales y humanos suficientes, mandato por periodos mínimos de dos años renovables, comunicación de datos de contacto a la autoridad de control del Estado miembro en cuestión y al público en general (cualquier interesado tendrá derecho a contactar con el DPO) para el debido cumplimiento de una de sus funciones, la de atender consultas, quejas y reclamaciones de los afectados.

Parte III

Principales retos de la Economía de los Datos

7

Titularidad y residencia de los datos

La Estrategia Europea del Mercado Único Digital consta de tres pilares: (i) mejorar el acceso de los consumidores y las empresas a los bienes y servicios digitales en toda Europa; (ii) crear las condiciones adecuadas y garantizar la igualdad de condiciones para que las redes digitales y los servicios innovadores puedan prosperar, y (iii) maximizar el potencial de crecimiento de la Economía Digital, que se sustenta en la propuesta de la Unión Europea de crear derechos para los productores de datos.

En este contexto, el *big data* presenta una doble vertiente. Por un lado, es el conjunto de herramientas de recolección, tratamiento y análisis de datos, protegidas por el derecho de propiedad intelectual mediante patentes o derechos de autor. Por otro, los datos, generados por diversos agentes, cuyo acceso ha de estar protegido en favor de la innovación, una protección que genera cuestiones de fondo tales como el derecho de exclusividad para el productor de los datos, acceso vía concesión de licencias, acceso remunerado a datos anonimizados, aplicación de principios de asimetría en la contratación a partir de la inclusión de obligaciones de transparencia, control de cláusulas abusivas, etc.

Colaboración 6. Sobre la titularidad y la residencia de los datos

Borja Adsuara, jurista y experto en estrategia digital

Para hablar de la Economía de los Datos, basada en la tecnología *big data*, y de sus principales retos —entre los que se encuentra, sin duda, su regulación—, tenemos que hablar antes de qué es un dato y de cuántos tipos de datos hay. Luego hablaremos de la titularidad y la «residencia» (o la «tenencia») de los datos, así como de los derechos inherentes a estas dos relaciones y/o situaciones jurídicas, tanto los morales (o personalísimos), como los de explotación (o económicos). Por último, haremos referencia a las cosas que se pueden hacer con los datos: la protección (de los datos personales, protegidos) frente a la «transparencia» y, sobre todo, la apertura (*open data*) y reutilización de datos (seudonimizados¹).

A efectos pedagógicos, porque en esta materia hay gran confusión conceptual, plantearemos estos temas en forma de «preguntas más frecuentes» (FAQ), cuyas respuestas deberían orientar la regulación de la Economía de los Datos.

Concepto y clases de datos

¿Qué es un «dato»?

El término «dato» proviene del participio pasado del verbo latino *dare* (dar) y significa algo que nos viene «dado», como la fecha (en inglés, *date* tiene el mismo origen etimológico).

1. Seudonimización: el tratamiento de datos personales de manera tal que ya no puedan atribuirse a un interesado sin utilizar información adicional, siempre que dicha información adicional figure por separado y esté sujeta a medidas técnicas y organizativas destinadas a garantizar que los datos personales no se atribuyan a una persona física identificada o identificable. Nuevo concepto jurídico en las definiciones dispuestas en el artículo 4 del Reglamento GDPR.

- Diferencia con una «obra»: Un dato se diferencia de una «obra» en que esta es producto de la inteligencia humana. Por ello, una obra está protegida por la Ley de Propiedad Intelectual, mientras que un dato, en su caso, por la Ley Orgánica de Protección de Datos.
- Tipos de datos: Hay muchos posibles criterios de clasificación de los datos, pero el que más nos interesa, a efectos jurídicos, es el que los divide en datos de carácter personal y datos que no son de carácter personal.
 - ¿Qué es un dato personal?: Según nuestra LOPD es «cualquier información *concerniente* a personas físicas identificadas o identificables». Definición muy similar a la del nuevo RGPD de la Unión Europea: «Toda información *sobre* una persona física identificada o identificable».
 - ¿Son datos personales los datos del Internet de las Cosas?: Algunos piensan que el Internet de las Cosas no está sujeto a la LOPD-RGPD porque no son datos de personas sino de cosas, pero lo importante no es cómo se recogen y transmiten esos datos, sino si *conciernen a* o son *sobre* personas.

Titularidad y residencia de los datos

- ¿De quién son los datos personales?: Los datos son de «su» titular: la persona física a la que le *conciernen* o le *afectan* (en su honor, su intimidad o su propia imagen), pero esa relación (de *suidad*) no es una relación o un derecho de «propiedad», sino un «derecho personalísimo», como el derecho al honor, a la intimidad y a la propia imagen, derechos a los que está vinculada la protección de datos, por el artículo 18 de nuestra Constitución.
- ¿Qué es lo que en realidad se protege?: En realidad, los datos no son de nadie. Por ejemplo: si Juan ha nacido en España, el dato España no es «propiedad» de Juan, sino que es un dato compartido con todos los que han nacido en nuestro país. Lo que realmente «es» de una persona (y se protege) es su honor, su intimidad y su propia imagen, que pueden verse afectados o concernidos por dichos datos, al proporcionar información sobre ella.
- ¿Titular, afectado o interesado?: En la LOPD se habla indistintamente de «titular», «afectado» o «interesado» en los datos personales, cuando, en rigor, son cosas distintas. El RGPD generaliza, en su versión en español, el término «interesado», cuando, sin embargo, en la versión en inglés se habla de *data subject* (sujeto o titular de los datos) y en la versión en francés, de *personne concernée* (persona concernida), que es más correcto.
- ¿Qué es la «residencia» o «tenencia» de los datos?: Una cosa es de quién «son» o a quién «afectan» los datos y otra cosa diferente es quién los tiene y los trata o gestiona. También se puede hablar de una soberanía o nacionalidad de los datos frente a la residencia de los datos en un país distinto. En todo caso, la titularidad no se pierde. Los datos no son de quienes los tienen; más bien, lo que tienen estos es una serie de obligaciones respecto de aquellos.
- ¿Derechos de explotación o derechos morales?: Igual que sucede en propiedad intelectual, sobre el derecho al honor, la intimidad y la propia imagen de una per-

sona y sobre los datos que le conciernen o afectan, pueden existir «derechos de explotación», de contenido económico o patrimonial, que pueden cederse a un tercero (con un consentimiento expreso y específico), pero nunca se renuncia a los «derechos morales», de contenido más «espiritual». Ejemplo: Sobre una misma fotografía puede haber un derecho de propiedad material (del poseedor), un derecho de propiedad intelectual (del autor) y un derecho al honor, a la intimidad y a la propia imagen (de quienes aparecen en ella). Lo mismo pasa con un dato (la imagen de una persona es un dato personal): puede tenerlo una persona, pero no puede hacer lo que quiera con él, si afecta a otra persona.

Economía de los Datos

- ¿Qué se puede hacer con los datos?: Como dijimos al principio, con los datos se pueden hacer dos cosas: «cerrarlos» o protegerlos, si son «datos personales» (que afectan a personas identificadas o identificables), o «abrirlos» (*open data*) y reutilizarlos, si están seudonimizados.
- ¿Hay que proteger siempre todos los datos personales?: No hay que proteger siempre todos los datos personales, sino que se pueden usar —y hasta explotar económicamente— siempre que haya un consentimiento del titular, un «interés público» o, incluso, un «interés legítimo» de una empresa.
- ¿Qué «modelos de negocio» hay en la Economía de los Datos?: Básicamente hay dos: 1) los que utilizan los datos de clientes de servicios de pago para prestarles mejor dichos servicios y 2) los que captan los datos de clientes de servicios gratuitos (como moneda de pago) para luego cedérselos a terceros.
- ¿Cómo pueden favorecer los datos de las Administraciones públicas la Economía Digital?: Los datos personales que tienen y tratan las Administraciones públicas no son «suyos», sino de los ciudadanos. Ellas solo los administran, con la obligación de custodia y secreto, pero pueden y deben abrirlos (seudonimizados), para que sean reutilizados.
- ¿Qué beneficios obtiene la sociedad de la apertura de datos?: Aparte de la creación de empresas, puestos de trabajo y riqueza (que revierte en las Administraciones públicas a través de los impuestos), la apertura y la reutilización de los datos de las Administraciones públicas favorece proyectos sociales: de salud, seguridad, medio ambiente, etc.

El consentimiento, el interés legítimo y el interés público

La explotación y la Economía de los Datos se basan en que el tratamiento sea lícito. Y, para ello, cuando son datos personales, hay tres vías: el consentimiento del titular, el «interés legítimo» de una empresa o el «interés general» de las Administraciones públicas.

Hasta ahora, en la LOPD, quedaba muy claro que el consentimiento del titular (afectado o concernido por los datos) era la regla general, y el «interés general» de las Administraciones públicas o el «interés legítimo» de las empresas, eran las excepciones.

Sin embargo, el nuevo RGPD pone al mismo nivel estas tres causas de licitud y, además, elimina la condición —que existía en la LOPD— de que los datos solo podían captarse, sin consentimiento, de ciertas «fuentes accesibles al público».

A estas alturas, todavía no sabemos hasta dónde llegará la interpretación del «interés legítimo» de una empresa para poder tratar los datos de una persona sin su consentimiento, aspecto clave para la explotación y la Economía de los Datos.

El reto del marco regulatorio

Nos encontramos actualmente en un momento de inseguridad jurídica respecto al marco regulatorio, no solo de la protección de los datos de carácter personal, sino, sobre todo, de la gestión y explotación de dichos datos.

El 25 de mayo de 2016 entró en vigor el Reglamento General de Protección de Datos de la Unión Europea, pero no se aplicará (su régimen sancionador) hasta el 25 de mayo de 2018, por lo que hasta entonces seguirá vigente la LOPD.

No obstante, ya está en tramitación parlamentaria una modificación de la LOPD para adaptarse al RGPD y, al tiempo, una nueva directiva de privacidad en línea, que, seguramente, se «solapará», en alguna medida, con el RGPD.

Pero el marco regulatorio de los datos no se agota con este «bloque jurídico» de reglamentos (RGPD) y directivas europeas (privacidad en línea) y leyes (LOPD) y reglamentos (de desarrollo) nacionales, referidos a la protección de los datos. También están sujetos a la Directiva Europea y las siguientes leyes españolas: Ley de Reutilización de Información del Sector Público, Ley de Transparencia y Buen Gobierno y Ley de Secretos Oficiales.

Y no se deben olvidar el artículo 12 de la Declaración Universal de Derechos Humanos (ONU), los artículos 7 y 8 de la Carta de los Derechos Fundamentales de la Unión Europea, así como el artículo 18 de nuestra Constitución (en especial, el apartado 4). Además, se debe tener en cuenta que la protección de los datos personales, al igual que la del honor, la intimidad y la propia imagen, no solo se hace por vía administrativa, sino también por vía civil (Ley 1/1982) e, incluso, penal (Títulos X y XI del Código Penal), es decir, hay que contemplar el ciclo de vida entero del dato o la gestión del dato; no solo su protección, sino también su apertura y reutilización (y su explotación), con todas las garantías oportunas, porque en ello se basa la Economía de los Datos.

Hay que huir de la tentación de meter los datos en una caja fuerte e impedir su tratamiento y circulación, porque Internet y la Economía Digital se basan en ellos y, si se impide su circulación, se impide también el desarrollo de ambas. En este sentido, conviene recordar que el nombre completo del nuevo RGPD es el siguiente: «Reglamento relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos». Y el artículo 1.3 dice: «La libre circulación de los datos personales en la Unión no podrá ser restringida ni prohibida por motivos relacionados con la protección de las personas físicas en lo que respecta al tratamiento de datos personales».

La Economía de los Datos podrá desarrollarse en la medida en que se tenga claro que lo importante es la libre circulación de los datos. Con todas las garantías para los derechos de sus titulares, pero sin caer en la tentación de «encerrarlos».

Bibliografía del autor sobre la gestión de datos

El consentimiento. Capítulo X, en *Reglamento general de protección de datos: un nuevo modelo europeo de protección de datos*, José Luis Piñar Mañas (dir.), Reus, Madrid, 2016, pp. 151-170.

2014

- De la protección de datos a la prostitución de datos: 14/1/2014, elconfidencial.com
- *Big Open Data*: la boda entre el *big data* y el *open data*: 19/4/2014, elconfidencial.com
- Open Gata versus Gato Encerrado: 18/9/2014, elconfidencial.com

2015

- Cinco preguntas sobre *big data*, privacidad y el nuevo DNI electrónico: 20/1/2015, elconfidencial.com
- Santa Privacidad, virgen y mártir: 30/1/2015, elconfidencial.com
- Donar tus datos es tan solidario como donar tus órganos: 7/4/2015, elconfidencial.com
- ¿Crees que la información que manejan las Administraciones públicas te pertenece?: 4/5/2015, elconfidencial.com
- En el negocio de la banca digital, el nuevo dinero son los datos: 26/07/2015, elconfidencial.com

2016

- ¿Qué narices es un «delegado de Protección de Datos»? 7 preguntas clave: 25/4/2016
- ¿De quién son los datos?: 6/9/2016, elpais.com
- La sintaxis de los datos: 27/10/2016, elpais.com
- Hacia una política basada en datos: 14/11/2016, elpais.com

2017

- 20 millones de razones para que te preocupes de la protección de datos: 29/1/2017, bez.es
- Amo al Aura, pero esperaré hasta el consentimiento explícito: 4 dudas acerca de la 4.ª plataforma de Telefónica: 28/2/2017, elpais.com
- Los conceptos «interesado» e «interés» en el RGPD: 17/05/2017, lenguajejuridico.com
- Sexo y protección de datos: 18/6/2017, bez.es

8

Privacidad, soberanía, seguridad y transparencia de los datos

8.1

Soluciones a la privacidad y seguridad de los datos personales

147

8.2

Soberanía de los datos: una preocupación global

155

8.3

Seguridad de los datos

160

8.4

Transparencia de los datos

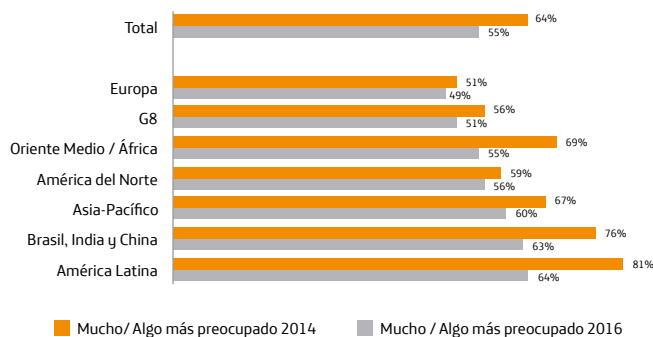
163

La sociedad aún no es plenamente consciente de la infinidad de información que revela diariamente, dejando un rastro de información que escapa a su control (por ejemplo, mediante la actividad en redes sociales, se pueden crear perfiles de usuarios según datos demográficos y, una vez procesados, pueden ser usados para generar publicidad o recomendaciones afines). Así lo recoge el último Special Eurobarometer on Data Protection, publicado en 2015 por la Comisión Europea, que concluye que el 36% de los españoles considera no tener ningún tipo de control sobre la información que ha facilitado por Internet, medido el mismo como la capacidad de corregir, cambiar o borrar dicha información a voluntad.

El estudio 2017 CIGI-Ipsos Global Survey on Internet Security and Trust, impulsado por el Centre for International Governance Innovation (CIGI) y ejecutado por Ipsos, entrevistó a 24.225 usuarios de Internet de 24 países¹ del 23 de diciembre de 2016 al 21 de marzo de 2017. Su objetivo fue apoyar el trabajo de la Global Commission on Internet Governance (GCIG), iniciativa de CIGI y Chatham House para articular y avanzar en la visión estratégica del futuro de la gobernanza de Internet.

Fig. 8.1

Grado de preocupación sobre la privacidad online (2016 versus 2014)



Fuente: 2017 CIGI-Ipsos Global Survey on Internet Security and Trust.

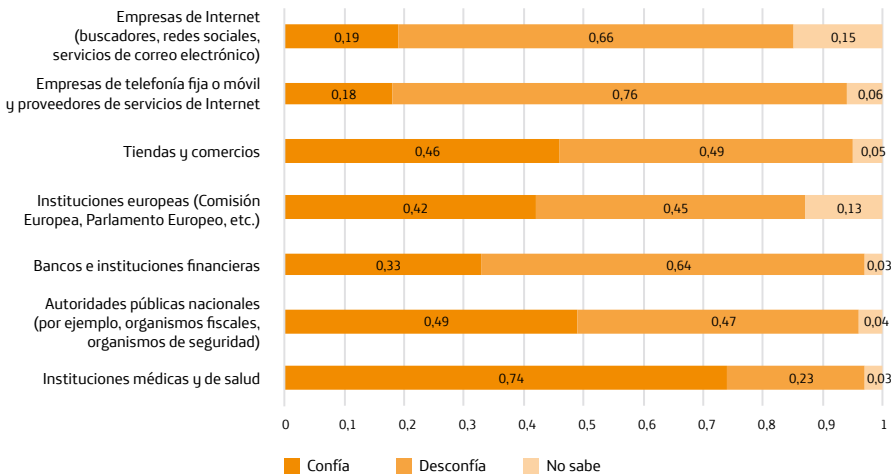
Según los resultados del estudio, se concluye que la mayoría de los internautas están más preocupados que hace dos años acerca de su privacidad online, especialmente en América Latina y Brasil, India y China. En el resto de regiones, el ritmo de crecimiento de la preocupación se atenúa año tras año. Por tanto, la gran mayoría de la población está preocupada por la información de carácter personal y sensible recopilada por Internet, comprada y vendida. Dicho esto, solo una minoría afirma que esta preocupación se traduce en un cambio de comportamiento online.

1. Australia, Brasil, Canadá, China, Egipto, Francia, Alemania, Hong Kong, India, Indonesia, Italia, Japón, Kenia, México, Nigeria, Pakistán, Polonia, Corea del Sur, Sudáfrica, Suecia, Túnez, Turquía, Reino Unido y Estados Unidos.

La confianza en las autoridades competentes es también mejorable: en España, el 47% de la población no confía en las autoridades públicas nacionales para proteger sus datos personales, y el 45% no lo hace en las autoridades europeas.

Fig. 8.2

Confianza de los ciudadanos españoles en distintas autoridades, públicas y privadas, para proteger su información personal (2015)



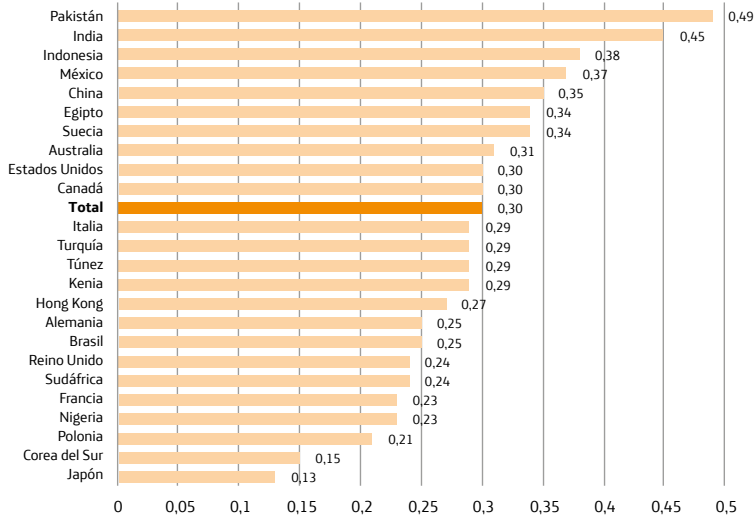
Fuente: Afi, a partir de Special Eurobarometer on Data Protection, julio de 2015, Comisión Europea.

A nivel global, según el CIGI-Ipsos Global Survey on Internet Security and Trust de 2016, el 70% de la población mundial manifestaba que los Gobiernos no hacen lo suficiente para proteger los datos personales que poseen las entidades privadas.

En la generación de los *millennials*, la confianza es sensiblemente superior que la declarada por otros grupos etarios de acuerdo a diversas encuestas efectuadas recientemente, como la realizada por Gallup en Estados Unidos, motivado fundamentalmente por su condición de nativos digitales y mayor intensidad de uso y participación activa en la Economía Digital en todos los ámbitos de la vida cotidiana.

Fig. 8.3

Población que considera suficiente la capacidad de los Gobiernos para proteger los datos personales que poseen entidades privadas (% sobre el total, según países)



Fuente: Afi, a partir de datos del Centre for International Governance Innovation (CIGI).

8.1

Soluciones a la privacidad y seguridad de los datos personales

Son múltiples las modalidades de soluciones para garantizar la privacidad de los datos personales en el entorno de la Economía de los Datos, cuya responsabilidad recae en los distintos agentes presentes en el ciclo de vida de los datos.

- Soluciones personales, centradas en la responsabilidad de autoprotección del propio usuario de dispositivos y canales de comunicación generadores de datos de carácter personal, privado o sensible. El ciudadano-consumidor habrá de equilibrar su exposición voluntaria e involuntaria a la captura de datos por los diferentes medios desplegados en su entorno, exposición facilitada por el equipamiento y el uso de los distintos dispositivos conectados a su disposición, con medidas básicas de protección de la privacidad. La primera de ellas es el conocimiento de los términos y las condiciones de uso y de privacidad de los datos que de forma constante son invitados a aceptar para hacer uso de múltiples aplicaciones y servicios digitales.

Los avisos de privacidad establecen los términos en los que las empresas hacen uso de la información recolectada, de modo tal que, si hacen un uso distinto al señalado en su aviso de privacidad, pueden ser sujetas a sanciones por parte de la autoridad. Todas las páginas web y aplicaciones que solicitan información personal lo deben tener.

La adecuada puesta en práctica de autoprotección precisa formar debidamente a los usuarios para mejorar su capacidad de comprensión en relación con las medidas deseables de protección de la privacidad y seguridad.

- Soluciones de mercado, entre las que destacan la autorregulación a la que se comprometen las empresas, ya sea a título individual o sectorial, así como la consideración de que el establecimiento de principios, políticas y prácticas que mejoren las garantías de privacidad de los datos de los usuarios son ventajas competitivas con respecto a otras empresas que no cuentan con dichas mejoras.
- Soluciones tecnológicas o *privacy enhancing technologies* (PET), definidas por la Comisión Europea como «aquellos sistemas coherentes de medidas TIC que protegen la privacidad mediante la eliminación o reducción de datos de carácter personal o mediante la prevención de tratamientos innecesarios o no deseados de datos personales, sin perder la funcionalidad del sistema de información».

Incluyen herramientas como la exigencia de uso de contraseñas y controles de acceso robustos; cortafuegos, comprobación de actualización de antivirus y otros *malware*; certificados de seguridad; auditorías; aislamiento de recursos no confiables; deshabilitación de funcionalidades innecesarias; aislamiento de componentes sensibles de la red; aseguramiento de copias de respaldo completas y mecanismos de recuperación de información; instalación automática de parches actualizados de seguridad, bloqueadores (de *cookies*, de sitios de Internet, etc.), y servicios para optar por no permanecer en bases de datos gestionadas por intermediarios. etc.

Las PET, por lo general, buscan proteger la información personal identificable (PII, por sus siglas en inglés) en sitios web, *smartphones* y otros dispositivos, y tienen menor aplicación en tecnologías como cámaras, lectores de tarjetas, sensores, etc., en cuyos casos son útiles otros enfoques tales como el control de revelación estadística (*statistical disclosure control*, SDC), la recuperación de información privada (*private information retrieval*, PIR) y la minería de datos para la preservación de la privacidad (*privacy preserving data mining*, PPDM), que buscan proteger la confidencialidad en el análisis de datos, la difusión de bases de datos públicas, así como en el minado de datos.

- Otras soluciones complementarias a las anteriores son las siguientes:
 - Principios y prácticas justas con la información (*fair information practice principles*, FIPPS), tales como el control individual, la transparencia,

el respeto por el contexto, la seguridad, el acceso y precisión, la captura restringida y la responsabilidad.

- Privacidad por diseño (*privacy by design*, PbD), entendida como la práctica operativa mediante la cual todos los datos son privados por defecto a menos que el usuario autorice expresamente lo contrario. Así, la privacidad forma parte de las especificaciones del diseño, de las prácticas de negocios, de los entornos físicos e infraestructuras de sistemas y aplicaciones. La privacidad por diseño es el enfoque regulatorio defendido por la Unión Europea, la Fair Trade Commission (FTC) de Estados Unidos y numerosas autoridades nacionales. Consta de siete principios:
 1. *Proactivo, no reactivo; preventivo no correctivo*: PbD no espera a que los riesgos se materialicen, ni ofrece remedios para resolver infracciones de privacidad una vez que ya ocurrieron; su finalidad es prevenir que ocurran.
 2. *Privacidad como la configuración predeterminada*: no se requiere acción alguna de parte de la persona para proteger la privacidad; está interconstruida en el sistema, como una configuración predeterminada.
 3. *Privacidad incrustada en el diseño*: la privacidad es parte integral del sistema, sin disminuir su funcionalidad.
 4. *Funcionalidad total*: acomodar todos los intereses y objetivos legítimos evitando falsas dualidades como privacidad versus seguridad.
 5. *Seguridad extremo a extremo (end-to-end)*: protección de ciclo de vida completo del dato.
 6. *Visibilidad y transparencia*: asegura a todos los involucrados que cualquiera que sea la práctica de negocios o tecnología involucrada, opera de acuerdo a las promesas y objetivos declarados, sujeta a verificación independiente.
 7. *Respeto por la privacidad de los usuarios*: mantener al usuario en el centro de las prioridades.
- Soluciones de gobernabilidad y gestión para que todas las funciones de datos se realicen del modo más eficiente, asegurando que los datos cumplan con las demandas, el cumplimiento de normativas y la preservación de la privacidad.

Tabla 8.1 Estrategias de privacidad

Fase <i>big data</i>	Estrategia	Implementación
Adquisición y recolección	Minimizar	Seleccionar antes de adquirir Evaluación de Impacto de la Protección de Datos (EIPD)
	Agregar	Anonimización en la fuente de origen
	Ocultar	Herramientas de cifrado Herramientas de enmascaramiento de datos
	Informar	Transparencia – comunicación al interesado
	Controlar	Mecanismos para recabar consentimiento
Análisis y validación	Agregar	Técnicas de anonimización
	Ocultar	Herramientas de cifrado
Almacenamiento	Ocultar	Herramientas de cifrado Mecanismos de autenticación y control de acceso
	Separar	Almacenamiento distribuido / descentralizado
Explotación	Agregar	Técnicas de anonimización
Todas las fases	Cumplir / demostrar	Definición de políticas Trazabilidad de las acciones Herramientas de cumplimiento

Fuente: AEPD.

Anonimización de los datos

En cada etapa del ciclo de vida del dato existen potenciales focos de riesgo de vulnerabilidad a la privacidad, como el que se podría dar mediante la combinación de distintos conjuntos de datos que no han atravesado un adecuado proceso de anonimización², y que podrían permitir la reidentificación de los individuos. La anonimización tiene muchísima relevancia en el contexto *big data*.

Los principales riesgos inherentes a la posibilidad potencial de identificación de un dato son los siguientes:

- singularizar, es decir, la posibilidad de aislar datos que identifican a un individuo en un conjunto;
- enlazar, es decir, la capacidad de vincular, al menos, dos datos referentes al mismo interesado o grupo de interesados (ya sea en la misma base de datos o en dos bases de datos diferentes), e
- inferir, es decir, la posibilidad de deducir, con una probabilidad significativa, el valor de un atributo en un conjunto de atributos.

Pero existen técnicas que minimizan dichos riesgos y principalmente son las siguientes:

2. Para que un dato sea verdaderamente anónimo ha de ser completamente irreversible su identificación.

- Asignación al azar: alterando la veracidad de los datos con el fin de eliminar el vínculo entre los datos y el individuo, se encuentran técnicas como:
 - la adición del ruido: modificación de atributos en el conjunto de datos reduciendo su precisión pero conservando la distribución;
 - la permutación: desordenar los valores de los atributos para que queden artificialmente vinculados a diferentes titulares de los datos, y
 - la privacidad diferencial, cuando se retira por lo menos una información que afecta al titular.
- Generalización: disgregando los atributos de los interesados mediante la modificación de la escala respectiva o de orden de magnitud, mediante las técnicas *deaggregation* y *k-anonymity* que buscan impedir que un interesado sea señalado agrupándolo con, al menos, *k* otros individuos.

Colaboración 7. El derecho a la protección de datos

Cecilia Álvarez, presidenta de la Asociación Profesional Española de Privacidad (APEP)

El derecho a la protección de datos es un derecho fundamental de cuarta generación, que cobra su mayor importancia cuanto más se desarrolla la Sociedad Digital. La Sociedad Digital genera, necesita y analiza datos a gran escala y bajo coste, muchos de los cuales son «personales» porque se refieren a una persona física o permiten distinguirla de un conjunto. La utilidad de los datos personales se proyecta tanto en relación con un individuo determinado (por ejemplo, los contenidos y servicios personalizados de todo tipo, bancarios, médicos, etc.) como en la identificación de las tendencias o patrones de conducta agregados (donde no se busca a un individuo específico, sino comprender un universo más amplio). En ambos casos, quién decide para qué se utilizan «mis» datos y cómo, puede tanto ayudar a desarrollar o amenazar mi dignidad y mi libertad.

Es habitual que se parta de la idea de que la utilización de «mis» datos requiere siempre mi consentimiento, como si la soberanía del dato recayera únicamente en nosotros. Sin embargo, el consentimiento tiene límites innegables en la Sociedad Digital. Además, el derecho moderno de protección de datos personales no se ha construido nunca ni debe hacerse únicamente sobre el consentimiento no solo por los límites de este sino también porque hay otros intereses en juego dignos de protección. Por ello, sin abandonar el consentimiento en todos los casos, quizá deberíamos utilizar otras herramientas para garantizar una protección efectiva o crear nuevas.

Los límites del consentimiento para garantizar una protección efectiva

Existen pocas instancias en las que el consentimiento pueda servir como garantía de protección efectiva. Primero, porque exige que leamos y comprendamos las políticas de privacidad. Parece que las generaciones actuales no leen ningún documento

(un triste hecho que demuestra la dejación de nuestra responsabilidad, pero ello no constituye una justificación) y no es menos cierto que las organizaciones redactan las políticas de privacidad para pasar el examen de las autoridades de protección de datos más que para las personas a las que están dirigidas. En todo caso, resulta difícil explicar en pocas palabras a los usuarios, a quién o a qué consiente en muchos supuestos del Internet de las Cosas, donde la complejidad es notoria, por los innumerables flujos de datos y actores involucrados.

Segundo, porque el consentimiento «real» debería ser «libre»: sin embargo, en la práctica, vivimos en un contexto de «lo tomas o lo dejas». Esta circunstancia no debería ser tan preocupante si se sancionaran las conductas abusivas. En efecto, el desarrollo de Internet le debe mucho a la utilización de términos y condiciones generales que no se negocian por su propia naturaleza, pero tienen su contrapunto en la sanción civil de nulidad (además de la sanción económica teórica) cuando se utilizan, en sede de consumidores y usuarios, términos y condiciones generales abusivos. Tampoco sería tan grave si valoráramos nuestra privacidad: como usuarios y como organizaciones. La elección de un producto o servicio no solo debe guiarse por su precio sino por el compromiso real de la organización con la protección del medio ambiente, con el bienestar de los animales, con la falta de explotación de la pobreza y... con la privacidad.

Tercero, el consentimiento no debería «sanar» una conducta ilegítima. Esto es, si la conducta es nociva para la dignidad del ser humano o su libertad, quizá no debiéramos dejar que se legitimara por el «consentimiento» sino prohibirla o rodearla de garantías.

Más allá del consentimiento

En la Unión Europea, la protección de los datos personales es un derecho fundamental. Como ocurre con otros derechos fundamentales de la personalidad, con el que guarda una cercana relación, como es el caso del derecho a la intimidad y a la propia imagen, el derecho a la protección de los datos personales no es absoluto. Es un derecho sobre el que cada uno de nosotros tiene cierta capacidad de decisión, pero esta no es absoluta porque hay otros intereses que deben ser tenidos en cuenta.

Por ello, las normas de la Unión Europea de protección de datos personales han tenido en cuenta al consentimiento como una de las posibles alternativas para legitimar un tratamiento, pero no la única. Esas causas de legitimación son diversas, pero todas ellas tienen en cuenta otros intereses con los que mi derecho fundamental debe convivir. En particular, destaca, por ejemplo, la necesidad de cumplir deberes legales o el desarrollo de un interés público (por ejemplo, pagar impuestos, la protección contra el blanqueo de capitales, la tutela judicial efectiva, la defensa de la competencia del mercado, etc.), la protección del interés vital de la persona o los actos propios de la persona (su voluntad de querer estar vinculado contractualmente con una entidad o el diagnóstico médico o los datos hechos manifiestamente públicos por la persona).

La garantía de protección en el primer caso es que los deberes legales, el reconocimiento de un interés público o la competencia de una Administración pública deben venir determinados por una ley aprobada en un Parlamento (y no un acto gubernamental discrecional, ya que no hay que olvidar que los derechos fundamentales nacieron como una limitación frente al poder del Estado). En el segundo caso, se entiende que la vida o integridad física de la persona debe prevalecer sobre este derecho. En el último caso, es un ejercicio de ser consecuente con las decisiones adoptadas por uno que afecta a derechos e intereses legítimos de otros: en la medida en que la relación contractual que libremente contraje siga viva (relación laboral, un contrato de préstamo o de arrendamiento, etc.), los datos necesarios deben seguir siendo tratados para dicha finalidad; de la misma forma que si decido publicar mis datos personales, debo atenerme a las consecuencias (con ciertos límites).

Adicionalmente a estos supuestos, en la Sociedad Digital, será particularmente necesario estar preparado para otra razón por la que el tratamiento de datos personales no deba estar siempre basado en mi consentimiento *ab initio*: se trata de la denominada razón del «interés legítimo». La razón del interés «legítimo» es probablemente la garantía más compleja y, por ello, la más efectiva, porque conlleva un análisis responsable de la convivencia de todos los intereses en juego (los de la organización, los de los individuos afectados, los de la sociedad en su conjunto) y qué salvaguardas al derecho fundamental deben incorporarse en su caso. Cuanto más se confunda el interés legítimo de una organización con un interés social, menos importancia tendrá la capacidad del individuo y viceversa. Por ello, la cesión de datos personales consustancial a la cesión de un derecho de crédito (donde la legislación civil ya ha decidido que no requiere el consentimiento del deudor porque haría inviable el sistema crediticio) no puede examinarse de la misma forma que el tratamiento de datos para remitirme publicidad (donde mi capacidad de decisión debe poder garantizarse), o para proteger la seguridad colectiva (donde la calidad de las bases de datos, la proporcionalidad de los supuestos de injerencia que debe ser transparentes y sujetos a control judicial son esenciales para garantizar la esencia del derecho fundamental).

¿Qué otras herramientas tiene el derecho fundamental en la Sociedad Digital?

Así las cosas, ¿cómo podemos garantizar el derecho fundamental a la protección de datos personales? Ya hemos hablado del análisis que conlleva la razón del interés legítimo que obliga a tener un análisis responsable y holístico, en vez de parapetarse en un consentimiento falaz. Para ello, nuevos profesionales deben ayudar a realizar este y otros análisis del impacto de la privacidad en las iniciativas de las organizaciones públicas y privadas. Ha nacido una nueva profesión, la de los profesionales de la privacidad, híbridos de juristas y tecnólogos.

Asimismo, no hemos mencionado aún que, con el Reglamento Europeo de Protección de Datos, contamos asimismo con el principio fundamental de la privacidad por diseño (PbD), una de las mejores garantías para que muchos procesos

tecnológicos donde intervengan actos de tratamiento de datos personales incorporen los principios fundamentales de protección de datos. Y hablamos de «muchos» pero no de «todos» porque esta obligación de privacidad por diseño no se ha impuesto (como debería) a los fabricantes de tecnología sino a los que se consideran «responsables» o «encargados de tratamiento». La seguridad técnica de los datos es sin duda uno de los elementos indispensables de la privacidad por diseño, pero la seguridad no agota las obligaciones que derivan de aplicar la privacidad por diseño. En efecto, no debemos olvidar en qué reside el carácter fundamental del derecho a la protección de datos personales. El derecho a la protección de datos sigue siendo, y más aún en la Sociedad Digital, del *big data* y del *machine learning*, una garantía de la libertad y dignidad. *Privacy by design* significa principalmente que el desarrollo tecnológico no debe hacernos olvidar que este está al servicio del hombre y no al revés. Debemos comprender los procesos técnicos que utilizamos para tomar decisiones sobre nosotros y nuestro entorno, de manera que podamos corregirlos y mejorarlos, sin caer en la «dictadura del dato» y sin desaprender, así como asumiendo la responsabilidad de nuestros actos, pues detrás de un dato, hay un ser humano.

En tercer lugar, hay que abordar el valor económico del dato personal. El que se trate de un derecho fundamental y, por ello, de un derecho inalienable, no significa que no se deba abordar su faceta patrimonial. Nadie parece alarmarse con la explotación por el individuo de su intimidad (en particular, ha sido moneda común ver a *celebrities* de toda naturaleza, incluidos deportistas de élite, negociar con su intimidad o sus derechos de imagen). Sin embargo, en marzo de 2017, el supervisor europeo de Protección de Datos, al afirmar que los datos personales no deben servir como divisa para «pagar» servicios online,³ pone el énfasis en una cuestión importante que es el aspecto patrimonial del dato para el individuo. En este sentido, siempre he visto una cierta similitud de los datos personales con los derechos de propiedad intelectual en la Unión Europea, coexistiendo una dimensión moral innegociable y una dimensión patrimonial que permite la explotación económica del derecho, protegiendo al autor por defecto y donde la ley establece límites al derecho exclusivo para dar cabida a otros intereses también dignos de protección. Sin embargo, en la protección de datos personales, no se atribuye un derecho de propiedad a nadie como tal, sino obligaciones a los poseedores de los datos (responsables o encargados de tratamiento, en la terminología técnica). Debemos abordar esta cuestión con responsabilidad y sin dogmatismos. Resulta sorprendente que la Comisión Europea se esté planteando cómo regular los derechos de propiedad únicamente sobre los datos «no personales». Y ello no tiene sentido por dos motivos: no solo porque muchos de los datos que considera «no

3. El Supervisor Europeo para la Protección de Datos (EDPS, por sus siglas en inglés) emitió una Opinión Pública en marzo de 2017 respaldando la iniciativa de la Comisión Europea de mejorar los derechos de los consumidores en relación a los aspectos concernientes a la contratación de servicios de contenidos digitales: https://edps.europa.eu/sites/edp/files/edpsweb_press_releases/edps-2017-3-digital_content_en.pdf

personales» (en particular, muchos *machine-to-machine generated data*) son datos personales, sino porque es innegable que los datos personales son un activo de negocio. La única manera de garantizar que el individuo obtenga una parte del «pastel» del valor que ayuda a generar es reconocer el aspecto patrimonial de los datos personales.

Conclusión

La Sociedad Digital nos da la oportunidad como personas de valorar nuestra privacidad y exigir su respeto a las organizaciones con las que tenemos contacto, lo que no debe confundirse con el consentimiento en todos los casos. Como personas objeto de protección, debemos ser responsables de nuestros actos propios, actuar en aquellos contextos donde el consentimiento tenga su plena razón de ser (por ejemplo, gestión dinámica y contextual de las preferencias de publicidad) y reflexionar sobre el impacto que para nosotros mismos y para otros puede tener la información que decidimos hacer pública.

El consentimiento tiene sus límites, por lo que la garantía efectiva de ese derecho requiere que a las organizaciones (privadas y públicas) se les exija su compromiso en esta materia (*accountability*), abrazando el principio de *privacy by design*, dedicando esfuerzos tanto a la seguridad técnica como al arte de explicar el valor añadido de lo que recibimos a cambio de los datos y tener una aproximación ética respecto del impacto del tratamiento de datos personales en los intereses en juego.

Los legisladores y las autoridades de control no deben olvidar dónde reside el valor de la protección de datos personales, que no es otra cosa que una garantía de la libertad y dignidad. Por ello, también ellos deben ser conscientes de que no es un derecho absoluto: se necesita un *test* permanente de los intereses legítimos y la gestión del riesgo. Asimismo, deben abordar el peligro de la deshumanización y la irresponsabilidad al que nos puede llevar el determinismo tecnológico, así como la necesidad de regular la propiedad de los datos personales.

8.2

Soberanía de los datos: una preocupación global

Las cuestiones relativas a la soberanía, la privacidad y la residencia de los datos son prioritarias para Gobiernos de todo el mundo. La definición de soberanía de los datos se basa en el concepto de que la información creada o almacenada de forma digital está sujeta a las leyes y regulaciones del país en que dicha información se encuentra localizada.



El gobierno de los datos y su adscripción a las jurisdicciones donde estos se generan —o donde radican las personas físicas, jurídicas o activos que los producen— es un aspecto crítico que impacta de pleno en el desarrollo de la Economía Digital. No en vano, las decisiones que en fechas recientes han sido adoptadas por distintos países en relación con la ubicación de sus centros de datos, tienen como motivación principal, si no única, la residencia y el gobierno de los datos.

Los principales temas de debate son los siguientes: (i) el cumplimiento normativo a nivel doméstico (confidencialidad, disponibilidad, retención, auditoría); (ii) la jurisdicción reclamada por terceros países; (iii) las fugas de datos (*data leakage*) y los riesgos de privacidad; (iv) la responsabilidad, el remedio, las garantías, los seguros, y (v) la consistencia legal entre jurisdicciones.

El diferente desarrollo normativo establecido por país ha creado un nuevo mapa global de soberanía de los datos, dividido entre los países neutrales —que muestran un posicionamiento abierto— y no neutrales, aquellos que consideran que la información de sus ciudadanos no debe ser almacenada ni procesada fuera de sus fronteras, estableciendo así una soberanía de datos.

Esta mayor o menor neutralidad se traduce, en términos prácticos, en la decisión de la ubicación de los centros de datos, hoy infraestructuras y servicios críticos y trasladados al ámbito interno del perímetro de la seguridad nacional. En julio de 2017, Apple Inc. anunció la creación de su primer centro de datos en China para cumplir con la regulación del país que exige que el almacenamiento de datos tenga lugar dentro de las fronteras domésticas. Dos años antes, el Gobierno ruso modificó las condiciones sobre la ubicación de los centros de datos de los proveedores de servicios digitales.

El exponencial crecimiento en el uso de servicios basados en Internet obliga a un paralelo crecimiento del tamaño de las infraestructuras de centros de datos que soportan estos servicios. Las compañías que los ofrecen enfrentan adicionalmente el reto de minimizar el impacto ambiental de sus operaciones, por su intensivo uso de energía.

Fig. 8.4

Ubicación de los centros de datos de Google

Fuente: www.google.com

El poder de almacenamiento en la nube radica en su potencial para ofrecer un proceso estandarizado y simplificado que elimina las fronteras geográficas y físicas. La innovación impulsada por la libertad de almacenamiento de datos en la nube radica en su capacidad para proporcionar dos atributos esenciales que proporcionan enorme flexibilidad: acceso en cualquier momento y desde cualquier dispositivo.

La amenaza del control regulatorio condiciona las capacidades de los servicios en la nube en el mundo; en una dinámica generadora de efectos dominó a partir de las respuestas de algunos países al establecimiento de medidas de seguridad y acciones de privacidad de otros, motivando una tendencia al almacenamiento y el tratamiento de los datos dentro de las fronteras domésticas: hoy no son pocos los países que exigen que los datos de sus ciudadanos se almacenen en el país en cuestión, y no en el de ubicación de la razón social de las empresas correspondientes. Esta tendencia se activó tras la promulgación de la Patriot Act por Estados Unidos tras los atentados del 11 de septiembre de 2001. Ello obliga a las empresas a duplicar sus datos para cumplir con las leyes de cada uno de los países en que operan, generando elevados costes por cumplimiento normativo, por las complicaciones de la minería de datos y por los mayores riesgos de dispersión entre varios proveedores de servicios de almacenamiento en la nube, con los que hay que verificar que cumplen asimismo con la regulación de soberanía y los *Service Legal Agreements* (SLA).

Cloud computing

Cloud computing, o servicios en la nube, es una de las tecnologías digitales consideradas importantes habilitadoras del aumento de la productividad y la mejora en

la provisión de servicios. Permite optimizar el uso de los recursos, así como crear modelos de negocio y estrategias de mercado en la Economía Digital.

Cloud computing presenta tres formas de contratación: (i) infraestructura como servicio (IaaS), que proporciona acceso a las características de redes, a los equipos y *software* dedicado y al espacio de almacenamiento de datos; (ii) plataforma como servicio (PaaS), que permite centrarse en la implementación y la administración de aplicaciones, y (iii) *software* como servicio (SaaS) o aplicaciones de usuario final.

En el marco de la Estrategia del Mercado Único Digital para Europa, el papel fundamental del *cloud computing* lo establecen la *European Cloud Initiative* y la Iniciativa para Construir una Economía de los Datos Europea (*Building a European Data Economy*, 2017), basada esta última en una propuesta legislativa para la regulación de la libre circulación de datos no personales y las restricciones a la localización de los datos. La propuesta contempla un enfoque de autorregulación a través de códigos de conducta para facilitar la portabilidad de los datos, así como para cambiar entre proveedores de servicios en la nube, consideradas estas condiciones necesarias para garantizar una Economía de los Datos competitiva en el marco de Mercado Único Digital (DSM). También aborda otros retos legales inherentes a las nuevas tecnologías sustentadas en datos, tales como el acceso y la transferencia de datos no personales generados de forma automática, la responsabilidad de los datos y la portabilidad, interoperabilidad y estándares.

Fig. 8.5

Mercado Único Digital: apoyando la nube en Europa



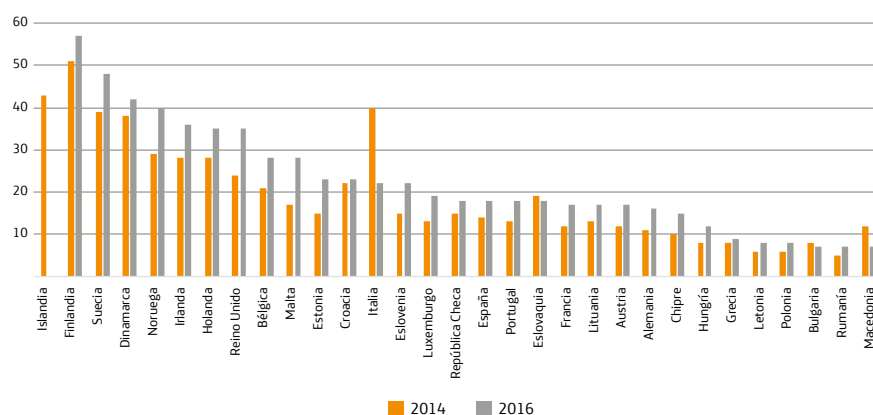
Fuente: Unión Europea.

La *European Cloud Initiative*, diseñada a partir de los resultados alcanzados en el marco de la *European Cloud Strategy* de 2012, y con insumos de la *European Cloud Partnership* (ECP) y el *Cloud Select Industry Group* (C-SIG), está centrada en una *European Open Science Cloud* (EOSC) —para almacenar, compartir y reutilizar de

forma abierta datos y resultados científicos— y en una Infraestructura de Datos Europea. Esta iniciativa está acompañada de la correspondiente a la digitalización de la industria (la denominada industria 4.0) —que incluye acciones para animar la adopción del *cloud computing*, especialmente entre las pymes—, con la definición de estándares como una de las áreas prioritarias en este particular. El paquete de conectividad a Internet pretende complementariamente proveer conectividad segura, ubicua y de alta calidad.

Fig. 8.6

Uso de *cloud computing* en la Unión Europea (% empresas), 2014 versus 2016



Fuente: Eurostat, [http://ec.europa.eu/eurostat/statistics-explained/index.php/File:V1_Use_of_cloud_computing_services,_2014_and_2016_\(%25_of_enterprises\).png](http://ec.europa.eu/eurostat/statistics-explained/index.php/File:V1_Use_of_cloud_computing_services,_2014_and_2016_(%25_of_enterprises).png)

El 21% de las empresas de la Unión Europea fue usuario en 2016 de servicios en la nube (19% en 2014), principalmente para alojar sus sistemas de correo electrónico y almacenamiento de archivos digitales. De estas, el 51% utilizó servicios en la nube avanzados (46% en 2014) —aplicaciones y *software* financiero y contable, CRM o capacidad de computación para determinadas aplicaciones de negocio—. La relación de uso de servidores públicos versus servidores de ámbito privado (infraestructuras de uso exclusivo) fue de 2:1 en 2016.

Estimaciones de la Comisión Europea sobre el desarrollo del *cloud computing* arrojan una dimensión para 2020 de cerca de 45.000 millones de euros (9.500 constatados en 2013).

8.3

Seguridad de los datos

La seguridad de los datos, la seguridad de la información o la seguridad informática es un aspecto esencial en organizaciones de cualquier tamaño y tipo. Se trata de la protección de datos contra accesos no autorizados, y para protegerlos de una posible corrupción durante todo su ciclo de vida. La seguridad de datos incluye conceptos como encriptación, *tokenización* y prácticas de gestión de claves que ayudan a proteger los datos en todas las aplicaciones y plataformas de una organización (marca, capital intelectual e información de clientes). En el mundo digital, el término ciberdefensa es el más extendido y se aplica a todos los elementos comunes de las organizaciones (personas, procesos y tecnología). El análisis de riesgo de datos incluye descubrirlo, identificarlo y clasificarlo, por lo que los administradores de datos pueden tomar medidas tácticas y estratégicas para asegurar que los datos sean seguros.

Son numerosos y crecientes los canales por los que transitan los datos. La protección y la monitorización de estos canales y de los datos que circulan por ellos es uno de los grandes retos de la seguridad en la Economía de los Datos. Los datos pueden encontrarse en distintos estados en función de su ubicación:

- Datos en reposo: residen en los sistemas de archivos, bases de datos y medios de almacenamiento tradicional, normalmente alojados en los centros de datos de las entidades.
- Datos en tránsito: salen al exterior de la entidad a través de redes públicas, generalmente Internet.
- Datos en el punto final (*endpoint*): almacenados en los terminales de los usuarios o en dispositivos de almacenamiento portátiles (USB, CD/DVD, discos duros externos, reproductores MP3, ordenadores portátiles o *smartphones*).

Evitar la pérdida de información sensible comienza por identificar la que es vital para la entidad, de manera que un buen punto de partida es el cumplimiento regulatorio y/o la protección de propiedad intelectual. En cuanto a las tecnologías que permiten proteger la información, destacan las siguientes:

- *Data loss prevention* (DLP): protege las fugas de información mediante un control de los repositorios y medios de transmisión dentro de la empresa, y
- *Enterprise data right management* (EDRM) o *information rights management* (IRM): para el control de acceso y uso de los archivos que hay que proteger, independientemente del lugar en el que se encuentren.

A punto de inaugurar la era del Internet de las Cosas (IoT), nos enfrentamos a nuevos riesgos a través de los dispositivos conectados que crean nuevas «conversaciones» (entre dispositivos, interfaces, infraestructuras privadas y la nube) suscep-

tibles de ser intervenidas, escuchadas o manipuladas por cibercriminales, quienes hacen uso de una nueva forma de extorsión digital, los ataques *ransomware*.

Prevenir, mitigar y reducir la vulnerabilidad de los sistemas de información y de datos es la misión de perfiles como el de los ingenieros de seguridad de datos, que protegen la red y otros activos físicos como servidores, computadoras y bases de datos. Para ello cuentan con distintas herramientas, tales como las siguientes:

- **Encriptación:** protege los datos y archivos almacenados o transmitidos a través de Internet. Los datos encriptados son ilegibles sin la clave de encriptación.
- **Sistemas de detección de intrusos de red (NIDS):** supervisan el tráfico de la red en busca de comportamientos ilícitos o anómalos.
- **Firewall:** *software* o *hardware* diseñado para bloquear el acceso a la red de usuarios no autorizados; líneas de defensa que evitan la interceptación de datos y bloquean el *malware* que intenta entrar en la red.
- **Análisis de vulnerabilidades:** identificación de amenazas potenciales, es decir, posibles «agujeros» en ordenadores, redes o infraestructuras de comunicaciones.
- **Pruebas de intrusión:** investigación de una red o sistema para detectar fallos mediante la ejecución de procesos que interrumpen los servidores, las aplicaciones, las redes e incluso los dispositivos de los usuarios finales para verificar si la intrusión es posible y por dónde se produce.
- **Información de seguridad y gestión de eventos (sistemas SIEM):** enfoque integral y centralizado que monitoriza cualquier lugar de la red (servidores, dispositivos de usuario o *software* de seguridad).
- **Conexiones cifradas y páginas seguras:** para crear canales de comunicación seguros existen los protocolos TCP/IP (con medidas de criptografía) y métodos de encriptación como *Secure Sockets Layer (SSL)* o *Transport Layer Security (TLS)*. Los protocolos HTTPS pueden ocultar y proteger los datos enviados y recibidos en los navegadores. Los *softwares anti-malware* y *anti-spyware* supervisan el tráfico de Internet entrante o el *malware (spyware, adware* o virus troyanos).
- **Detección de amenazas en punto final:** a pesar de que seguir buenas prácticas de seguridad —tener *software* antivirus, actualizado el sistema operativo y copias de seguridad de datos en la nube y en dispositivos locales—, previene ataques, los usuarios (empleados) y dispositivos con los que acceden a la red (teléfonos móviles, ordenadores portátiles o terminales punto de venta, TPV, móviles) son el eslabón más débil de la cadena de seguridad. Para minimizar riesgos, es preciso establecer distintos niveles de protección y de autorización de acceso a dispositivos de la red.
- **Prevención de pérdida de datos (*data leak prevention*, DLP):** medidas para asegurar que no se envían datos confidenciales desde la red. El *software* DLP supervisa la red y asegura que los usuarios finales autorizados copian o comparten datos que no deberían.

¿Seguridad *in-house* o en la nube?

La decisión de tercerizar el control de la seguridad de los datos a servicios en la nube es potestad de cada organización, que ha de valorar lo siguiente:

- Escala y capacidades de defensa: la experiencia, las capacidades técnicas necesarias y las economías de escala para administrar una nube son mayores que las que la mayoría de empresas individuales pueden tener.
- Ciclos de vida de desarrollo seguro: desde las soluciones de *software* hasta los sistemas de monitorización y los procesos de administración de la infraestructura están diseñados pensando en la seguridad de los datos.
- Auditoría continua: el *software* se actualiza a la última versión, se identifican y resuelven las anomalías en el rendimiento del sistema, y se cumplen los requisitos de cumplimiento de seguridad.
- Automatización y repetición: los niveles de automatización, normalización y auditoría garantizan que las firmas de virus y los parches de seguridad se actualicen rápidamente en toda la red.

Riesgos de incumplimiento

A pesar de la gran inversión en seguridad de la información y gobierno de datos, muchas entidades no terminan de comprender y proteger sus activos más valiosos —datos confidenciales y sensibles— y se encuentran expuestas a riesgos de incumplimiento, como observamos lamentablemente casi a diario.

El Reglamento General de Protección de Datos (GDPR) de la Unión Europea —ley a partir del 25 de mayo de 2018— impone penalizaciones muy significativas a las organizaciones que incumplan su mandato (hasta el 4% de su facturación global), incentivo que forzará a las entidades a entender sus riesgos de privacidad de datos, y tomar las medidas necesarias para reducirlos, ya que la responsabilidad de protección de la información es enorme, por las implicaciones y costes asociados a las brechas de seguridad de datos. Con estas condiciones, las organizaciones deben tener un conocimiento completo de sus datos confidenciales y sus riesgos para garantizar el cumplimiento de las políticas y leyes de privacidad.

La seguridad de los datos en la Unión Europea

Proteger las redes y los sistemas de información en la Unión Europea es esencial para dinamizar la Economía Digital. La ciberseguridad y ciberresiliencia fueron uno de los tres retos identificados en 2013 con motivo de la evaluación de medio término del Mercado Único Digital. El 13 de septiembre de 2017 la Comisión Europea adoptó un paquete de seguridad cibernética.

La Agencia Europea de Seguridad de las Redes y de la Información (ENISA, por sus siglas en inglés), creada en 2004 es el centro europeo especializado en ciberseguridad, ubicado en Grecia, ha sido objeto de una propuesta de reforma por parte de la Comisión Europea, para ampliar su mandato y fortalecer su capacidad de apoyo

a los Estados miembros. Entre otras novedades propuestas, la Comisión Europea considera que ENISA debe convertirse en agencia certificadora del futuro esquema de certificación en ciberseguridad; también debe apoyar en la implementación de la directiva relativa a las medidas destinadas a garantizar un elevado nivel común de seguridad de las redes y sistemas de información en la Unión (Directiva NIS), aprobada en julio de 2016, y en la gestión de crisis de ciberseguridad en la Unión Europea para la que la Comisión Europea ha presentado la Recomendación sobre la respuesta coordinada a los incidentes y crisis de ciberseguridad a gran escala.

La Unión Europea es firme defensora de que el derecho internacional y, en particular, la Carta de las Naciones Unidas, se aplique en el ciberespacio. También respalda normas voluntarias y principios de conducta responsable, y el desarrollo y aplicación de medidas de construcción de una confianza regional, tanto en el marco de la Organización para la Seguridad y Cooperación en Europa (OSCE) como en otras regiones.

El European Union Institute for Security Studies (EUISS) ha propuesto un nuevo marco de respuesta diplomática en el marco de la ciberdefensa (*Cyber Diplomacy Toolbox*), cuya implementación está teniendo lugar en los Estados miembros. Sobre cibercrimen, la Comisión Europea tiene previsto presentar a inicios de 2018 propuestas concretas que faciliten la transmisión transfronteriza de evidencia de tipo electrónico o digital.

8.4

Transparencia de los datos

Las ventajas de las nuevas técnicas de análisis de datos promueven la recolección continua y masiva de datos, así como su retención durante el máximo tiempo para una posible utilidad futura. Los datos en sí mismos adquieren valor por un potencial uso a futuro con el fin para el que fueron recabados y, por tanto, para el que los ciudadanos dieron su consentimiento expreso.

En este contexto, la Comisión Europea⁴ estima que el 64% de los españoles se declaran «preocupados» por la posibilidad que las instituciones y empresas privadas utilicen su información para un uso distinto del que dieron su consentimiento, sin comunicárselo.

La 36.^a Conferencia Internacional de Autoridades de Protección de Datos y Privacidad emitió una resolución sobre *big data*, en la que se señala que la protección proporcionada por los principios de derecho a la información y consentimiento es

4. Special Eurobarometer on Data Protection, julio de 2015.

más importante que nunca, por la enorme capacidad de recopilación de ingentes cantidades de datos personales con técnicas *big data*.

El RGPD busca garantizar que cada ciudadano tenga un mayor poder de disposición sobre sus datos de carácter personal, evitando que la protección suponga una restricción en el desarrollo del mercado interior de la Unión Europea. Por ello, el enfoque elegido no es el establecimiento de restricciones en el tratamiento de datos de carácter personal, sino la búsqueda de fórmulas que refuercen las garantías jurídicas que deben presidir dichos tratamientos de *big data*: control y transparencia informativa.

Para evitar que la exigencia general del consentimiento suponga un obstáculo a los desarrollos *big data*, se ha optado por dar máxima importancia al cumplimiento del derecho de información que todo titular de los datos tiene que tener sobre el tratamiento, uso y destino que se va a aplicar a los datos personales de su titularidad.

El concepto de transparencia está basado en la existencia de información suficiente sobre el tratamiento, al que tiene derecho el titular de los datos, conocimiento que puede ayudar a superar la necesidad del consentimiento e incluso del derecho de información del titular de los datos en la técnica *big data*. Según la AEPD, por tanto, cualquier uso previsto o futuro de los datos de carácter personal debe partir de la exigencia legal y moral del responsable del fichero y/o del tratamiento de actuar de forma transparente con el titular los datos.

Recuadro 8.1 La cuarta plataforma de Telefónica (AURA): modelo de relación con clientes basado en la transparencia

Telefónica presentó en el Mobile World Congress de 2017 su denominada cuarta plataforma (AURA), con la que los usuarios podrán gestionar su experiencia digital con la compañía y controlar de forma transparente y segura los datos que generan por el uso de sus productos y servicios, gracias a la inteligencia artificial.

Cuarta plataforma

- Inteligencia artificial

Tercera plataforma

- Productos y servicios

Segunda plataforma

- Sistemas unificados de tecnologías de la información

Primera plataforma

- Redes y activos fijos

Fuente: Afi, a partir de www.telefonica.es

AURA busca poner el control de los datos personales de los clientes de la compañía en sus propias manos: saber con qué datos cuenta y decidir si compartirlos o no con terceros.

Desde su generación en las tres primeras plataformas hasta su transformación en un valor único para cada cliente en la cuarta plataforma, los datos siguen un camino. Cada usuario contará con un *personal data space* —memoria donde se almacenará el rastro digital que dejan al utilizar los productos y servicios de Telefónica— que permitirá, si el cliente lo desea, personalizar su experiencia. Por ejemplo, con la función de *Timeline* el usuario podrá conocer los datos que se van generando al utilizar los productos y servicios de la compañía, de forma visual.

La cuarta plataforma permitirá la integración con otras tecnologías y servicios digitales, convirtiéndose así en una plataforma de innovación abierta a otros socios.

Por su dimensión, heterogeneidad y velocidad, la recolección de datos puede afectar a su calidad y, por tanto, a la calidad de las decisiones tomadas a partir de la misma. Los datos deben ser exactos, no excesivos y actualizados, recogiendo únicamente la información estrictamente necesaria para el objetivo en cuestión. El cumplimiento de dichos desafíos redundará en una mayor confianza de la sociedad, animándola a ser más proclive a participar en el desarrollo de nuevas y más eficientes iniciativas sustentadas en el *big data* y en *analytics*.

Recuadro 8.2 Resolución de la 36.^a Conferencia Internacional de Autoridades de Protección de Datos y Privacidad sobre el *big data*

Dado que el *big data* puede usarse en formas que generan una preocupación importante respecto a la privacidad de las personas y los derechos civiles y a las protecciones contra la discriminación y las vulneraciones al derecho a trato igual, la 36.^a Conferencia Internacional de Autoridades de Protección de Datos y Privacidad emitió una resolución que comprende los siguientes compromisos:

1. Respetar el principio de especificación de finalidad.
2. Limitar la cantidad de información recolectada y almacenada a un nivel que sea necesario para el propósito legítimo que pretende.
3. Obtener, cuando sea apropiado, el consentimiento válido del titular de los datos en relación con el uso de información personal para fines de análisis y de creación de perfiles.
4. Ser transparentes acerca de qué información se recolecta, cómo se procesa, con qué propósito será utilizada y si será transferida a terceros.
5. Dar a las personas acceso apropiado a los datos que han sido recolectados sobre ellas, y a la información y las decisiones que se han tomado con esos datos.
6. Ofrecer a las personas, cuando sea apropiado, el acceso a la información sobre los insu- mos principales y los criterios para la toma de decisiones (algoritmos) que se han utilizado como base para el desarrollo del perfil. La información debe presentarse en un formato claro y comprensible.
7. Llevar a cabo una evaluación de impacto en la privacidad.
8. Desarrollar y utilizar tecnologías del *big data* de acuerdo con los principios de la privaci- dad por diseño.
9. Considerar cuándo los datos anónimos mejorarán la protección de la privacidad. La ano- nimización puede ayudar a mitigar los riesgos para la privacidad asociados con el aná- lisis del *big data*, pero solo si la anonimización está diseñada y gestionada de manera apropiada. La solución óptima para anonimizar los datos debe decidirse caso por caso, posiblemente utilizando una combinación de técnicas.
10. Tener mucho cuidado y actuar cumpliendo la legislación aplicable en materia de protec- ción de datos, cuando se comparten o se publican conjuntos de datos con seudónimos o que pueden ser identificables indirectamente. El acceso debe ser limitado y controlado cuidadosamente si los datos contienen suficientes detalles, esto es, que puedan vincularse con otros conjuntos de datos o contienen datos personales.
11. Demostrar que las decisiones respecto al uso del *big data* son justas, transparentes y res- ponsables. Relacionado con el uso de datos para fines de creación de perfiles, tanto estos como los algoritmos en que están basados requieren una valoración continua. Este necesi- ta revisiones regulares para verificar si los resultados de la creación de perfiles son respon- sables, justos y éticos y si son compatibles y proporcionados con el propósito para el cual los perfiles son usados. Debe evitarse la injusticia con las personas debido a resultados completamente automatizados que arrojen un falso positivo o un falso negativo. Siempre debe estar disponible una valoración manual de resultados, con efectos significativos para los individuos.

9

La medición de la contribución de la Economía de los Datos en las cuentas nacionales

9.1

La metodología de estimación de la Economía de los Datos

172

9.2

La dimensión de la Economía de los Datos

176

Antes de justificar la importancia de los datos estadísticos, es necesario conocer qué se entiende por estadística. Según la Real Academia Española (RAE), la estadística es «la rama de la matemática que utiliza grandes conjuntos de datos numéricos para obtener inferencias basadas en el cálculo de probabilidades». Otra de las acepciones del término indica que la estadística es el «estudio de los datos cuantitativos de la población, de los recursos naturales e industriales, del tráfico o de cualquier otra manifestación de las sociedades humanas».

El Instituto Nacional de Estadística (INE), el organismo público e independiente que se encarga de hacer estadísticas en España, define a la estadística como la «ciencia del estudio de los datos asociados a un suceso o experimento que consiste en recogerlos, estudiarlos, analizarlos y sacar conclusiones para, posteriormente, proponer una ley que explique o describa el comportamiento de dicho suceso o experimento».

La historia de la estadística se remonta a las antiguas civilizaciones, cuando se recogía el número de personas que vivían en las ciudades, normalmente para organizar el pago de impuestos y el ejército. A lo largo de la historia, los Gobiernos y dirigentes de las distintas naciones han procurado disponer de otro tipo de datos, pero siempre con fines organizativos, ya que las estadísticas permitían hacer un seguimiento de las políticas públicas y dimensionar la provisión de los servicios públicos que ofrecían a la población.

Hoy en día, las estadísticas oficiales no se circunscriben al ámbito público, sino que abarcan a todo tipo de sectores y realidades socioeconómicas. La toma de decisiones a cualquier nivel, bien sea para innovar, progresar, conocer o diseñar políticas, se ha apoyado, con cada vez mayor frecuencia, sobre la evidencia empírica o, lo que es lo mismo, la evolución o tendencias observadas que reflejen los datos estadísticos. De hecho, un estudio¹ indica que «el valor de las estadísticas oficiales depende de su capacidad de crear conocimiento en toda la sociedad», y las considera un bien público, al reunir las características de calidad, imparcialidad y posibilidad de ser difundidas a todos los agentes de forma simultánea.

Por todo lo anterior, es necesario medir estadísticamente el tamaño de la Economía de los Datos, pues de ello dependerán las decisiones estratégicas más adecuadas para conocer en detalle su naturaleza y evolución, impulsar su desarrollo y corregir los problemas que puedan surgir en este proceso.

1. Giovannini, E.; Oliveira Martins, J., y Gamba, M. (2009): *Statistics, Knowledge and Governance*.

9.1

La metodología de estimación de la Economía de los Datos

El Producto Interior Bruto (PIB) es un indicador que se utiliza comúnmente para determinar el tamaño de una economía, y no es más que el conjunto de todos los bienes y servicios producidos en un país durante un periodo de tiempo determinado (generalmente un año). Por tanto, el PIB se calcula a partir de las operaciones de compraventa llevadas a cabo en el mercado que llevan asociado un determinado precio, es decir, se basa en lo que se conoce como la Economía de Mercado. Sin embargo, este enfoque convencional para la medición de la actividad económica, que ha servido para calcular el valor de los modelos de negocio tradicionales, no necesariamente puede valer para estimar el volumen de negocio que es capaz de generar la Economía de los Datos.

La revolución de los datos ha impregnado prácticamente todas las esferas de la vida cotidiana. Comunicarse o leer el periódico han pasado a realizarse de manera muy diferente a como se hacía hace algo más de dos décadas. La aparición de Internet y otros desarrollos tecnológicos han cambiado nuestro modo de consumir bienes y servicios, así como de relacionarnos. Es más, las personas ahora son capaces de proporcionarse a sí mismas la información que anteriormente adquirían en el mercado con una contraprestación económica. A modo de ejemplo, el consumidor de hoy puede detectar y reservar una habitación en un hotel o un vuelo directamente a través de Internet, sin necesidad de acudir a los servicios que ofrece una agencia de viajes, como se hacía en un pasado no muy lejano.

Este cambio de paradigma no implica una pérdida de valor de la actividad económica, quizá todo lo contrario, aunque parece inevitable pensar que, fruto de la libre competencia, sí supone un trasvase del volumen de la actividad desarrollada por los sectores económicos tradicionales a aquellos otros que emergen con motivo de la revolución digital. La dificultad de medir el valor que proporcionan estas plataformas al consumidor y su contribución al bienestar general de la economía ha abierto el debate sobre la capacidad de las estadísticas y metodologías empleadas hasta la fecha para captar, precisamente, las nuevas relaciones comerciales características de la Economía Digital, en general, y de la Economía de los Datos en particular.

Pero, por otro lado, parece impensable plantear que, precisamente ahora, en un momento en el que la tecnología ha progresado sobremedida y está a nuestra disposición para aprovecharla y medir adecuadamente todas las transacciones que se desarrollan en el mercado, sea complicado cuantificar la cantidad y valor de dichas operaciones, así como la parte que le corresponde a la Economía de los Datos. La tecnología debe considerarse un aliado para el perfeccionamiento de las estadísticas, no solo porque permite acceder a cantidades masivas de información recogida y transmitida digitalmente, sino también porque favorece el almacenamiento y el procesamiento, a veces automático, de la misma.

Esta revolución supone, por tanto, un replanteamiento de la manera en la que las estadísticas recogen la realidad socioeconómica que emerge con el desarrollo de la Economía de los Datos. Aunque esta revolución implique un importante reto metodológico y, en ocasiones, puede que incluso se considere difícil de superar, no sería la primera vez que se produciría una modificación de las estadísticas públicas de importante magnitud.

Así, por ejemplo, la reciente sustitución del Sistema Europeo de Cuentas Nacionales y Regionales,² que se extiende a todos los países miembros de la Unión Europea, se fundamentaba, entre otras cosas, en los siguientes cambios producidos en los últimos años:

- el creciente papel de las tecnologías de la información y las comunicaciones en los procesos productivos, y
- la disponibilidad de una mayor cantidad de información estadística más detallada, fiable y actualizada que permiten las nuevas tecnologías.

Este nuevo estándar metodológico implantado en 2010 no era más que una adaptación recomendada por la Comisión de Estadística de Naciones Unidas del Sistema de Cuentas Nacionales de 2008 (SCN-2008).³ Los primeros países en implantarlo fueron Estados Unidos, Australia y Canadá. Por tanto, los cambios metodológicos más relevantes no surgen en un único país o pequeño grupo de países, como tampoco se implementan de forma independiente. El objetivo de las estadísticas, además de ofrecer una fiel estimación del volumen de actividad económica que se registra en cada momento, es también garantizar la plena comparabilidad de la evolución reciente de las diferentes economías, de ahí la necesidad de que la metodología sea homogénea, es decir, esté armonizada, y requiera de un replanteamiento metodológico a escala global.

En este sentido, parece conveniente la creación de un marco de investigación que plantee la metodología asociada a la producción estadística oficial que surge como consecuencia del avance de la Economía de los Datos, y/o que mejore la metodología empleada hasta la fecha para adaptarla lo máximo posible a la realidad económica. Es importante también que el grupo de profesionales encargados de esta investigación, además de estar compuesto por los trabajadores de los organismos públicos, se amplíe al resto de la comunidad investigadora, de tal forma que maximice la rigurosidad de las metodologías propuestas y anticipe los resultados que puedan desprenderse de cada una de ellas, entre otras ventajas.

A nivel interno, como ya ha ocurrido en el pasado, el cambio metodológico que supone la Economía de los Datos exige la necesidad de adaptar las oficinas de estadística nacionales y, en concreto, la cualificación de sus recursos humanos, a través de programas de formación continua. El *big data* y las nuevas habilidades profesio-

2. De la metodología de 1995, conocida como SEC-1995, se ha pasado a la de 2010, comúnmente identificada con el acrónimo SEC-2010.

3. Sustituyó en su momento el SCN-1993.



nales requeridas para el aprovechamiento del mismo, como se analizaba con anterioridad, hacen imprescindible la recualificación de los trabajadores encargados de construir las futuras estadísticas públicas.

La OCDE⁴ ha sido uno de los primeros organismos internacionales en identificar ciertas lagunas en la medición de la Economía Digital en su amplio espectro, extensible también a la Economía de los Datos. Si bien las estadísticas existentes miden la difusión de las TIC, estas no tienen la misma capacidad para ir al paso de las nuevas tecnologías, en rápida evolución, y su utilización por parte de personas individuales y empresas. Es el uso y no tanto la penetración de la tecnología la que determina el valor que es capaz de generar la Economía Digital. Por ello, el estudio concluye que es necesario desarrollar una agenda de medición internacional orientada al futuro, que gire alrededor de las siguientes áreas:

- Mejorar la medición de la inversión en TIC y su vínculo con los resultados macroeconómicos, entre cuyas tareas se encuentran la medición y la valoración de los datos digitalizados al considerarlo un activo intangible, para estimar su contribución a la productividad y el desempeño empresarial.
- Definir y medir las competencias necesarias para la Economía Digital, a través del análisis de las estadísticas (públicas y privadas) sobre las habilidades profesionales, las ocupaciones y las ramas de actividad asociadas a la Economía de los Datos, que desemboque en un sistema armonizado de competencias digitales comparable a nivel mundial.

4. OCDE (2014): *Measuring the Digital Economy. A New Perspective*.

- Desarrollar métricas para realizar un seguimiento de cuestiones tan variadas como la seguridad, privacidad y protección al consumidor. Aunque ya se está avanzando en la armonización de estadísticas procedentes de los Equipos de Respuesta ante Incidencias de Seguridad (CSIRTs, del inglés Computer Security Incident Response Team),⁵ todavía queda mucho por hacer. A modo de ejemplo, los indicadores que se recomienda construir tienen que ver con el grado de incidencias de los ataques en la red, la construcción de alertas relativas a amenazas y vulnerabilidades que puede sufrir el usuario, etc.
- Fomentar la medición del impacto de la Economía Digital en la sociedad y de las TIC en general para alcanzar objetivos de carácter social (sector sanitario, medio ambiente, etc.). Entre las tareas que recomienda abordar se encuentra el desarrollo de encuestas a consumidores que mejoren su confianza y uso de la Economía Digital.
- Invertir en una infraestructura de información completa y de alta calidad para medir dichos impactos, a través de (i) la promoción del uso de los microdatos entre todos los agentes (empresas, organizaciones, trabajadores, hogares, etc.), (ii) la revisión y la recopilación de todas las estadísticas disponibles para aprovechar al máximo toda la información que proporcionan, y (iii) la mejora del acceso a los datos de la comunidad investigadora que garantice también la confidencialidad en el uso de la información.
- Desarrollar un marco estadístico de calidad adaptado para explotar Internet o cualquier otra fuente que ofrece la Economía de los Datos, como fuente de información complementaria a las que conocemos de carácter convencional. Entre las tareas que se propone desarrollar, se encuentran las siguientes: (i) el establecimiento de estándares estadísticos, como el muestreo; (ii) la elaboración de indicadores estadísticos (por ejemplo, los resultados de las búsquedas en Internet); (iii) el fomento de la colaboración entre empresas para el almacenamiento y procesamiento de la información, y (iv) el desarrollo del entorno regulatorio que vele por el correcto uso de los datos, etc.

Un ejemplo de una iniciativa para avanzar en la redefinición de la metodología de estimación de la actividad económica lo protagoniza Reino Unido, que es uno de los países más avanzados en la Economía de los Datos. El Gobierno británico ha encargado recientemente un estudio a la London School of Economics⁶ en el que evalúa las necesidades de las estadísticas actuales y futuras, así como la capacidad de su Oficina Nacional de Estadística (ONS) para satisfacerlas.

5. Se trata de un grupo de expertos responsable del desarrollo de medidas preventivas y reactivas ante incidencias de seguridad en los sistemas de información. Se encargan de estudiar el estado de seguridad global de redes y ordenadores y de proporcionar servicios de respuesta ante posibles incidentes.

6. London School of Economics (2016): Independent review of UK economic statistics: interim report.

Este estudio analiza, en primer lugar, los tres enfoques empleados hasta ahora para medir el valor de los productos y servicios digitales: (i) el gasto publicitario asociado, (ii) el valor del tiempo que se pasa en la red mientras que se consumen estos bienes y servicios y (iii) el tráfico de datos en la red. Aunque concluye que ninguna de las métricas es perfecta, sí es cierto que juntas dan una imagen de hasta qué punto la metodología convencional no capta bien el valor de estas actividades económicas. Por otro lado, el estudio también identifica a los datos como un nuevo factor de producción, como se considera al capital tecnológico o a los trabajadores. Según este estudio, cualquier acumulación de datos cuya vida útil supera el año debe registrarse y valorarse como una inversión en un activo fijo.

Aunque el estudio no llega a definir la metodología ideal para la estimación del valor de la Economía de los Datos, recomienda al Gobierno británico que cree un centro de excelencia que se encargue del seguimiento de las industrias que emergen en la era digital, realice un replanteamiento de la metodología de medición de la Economía de los Datos, y desarrolle valoraciones de nuevas empresas y/o sectores en particular que sirvan como casos prácticos.

En definitiva, comienzan a darse pasos en el replanteamiento de la forma de medir la Economía Digital, en general, y el mercado de los datos, en particular, con el objeto de considerar fielmente la realidad socioeconómica que emerge y el progreso al que contribuyen.

9.2

La dimensión de la Economía de los Datos

Un estudio desarrollado por IDC y Open Evidence⁷ para la Comisión Europea ha estimado el valor de mercado del dato para todos los países miembros de la Unión Europea. Su metodología se basa en las métricas comúnmente empleadas hasta la fecha para estimar el tamaño de un mercado, pues, como se indicaba con anterioridad, todavía no se ha diseñado un nuevo procedimiento. En particular, estima el valor de los productos y servicios relacionados con el dato que han sido intercambiados. Si bien estos productos y servicios se limitan a aquellos datos que han sido recogidos, procesados, elaborados y transmitidos a través de infraestructuras digitales de información o tecnologías digitales, no se agrega, como tampoco se hace para otros mercados, la facturación de las empresas digitales que operan en la Unión Europea y que intercambian este tipo de bienes y servicios, ya que esta magnitud tiene en cuenta las importaciones y excluye las exportaciones. El indicador que se pretende estimar debería con-

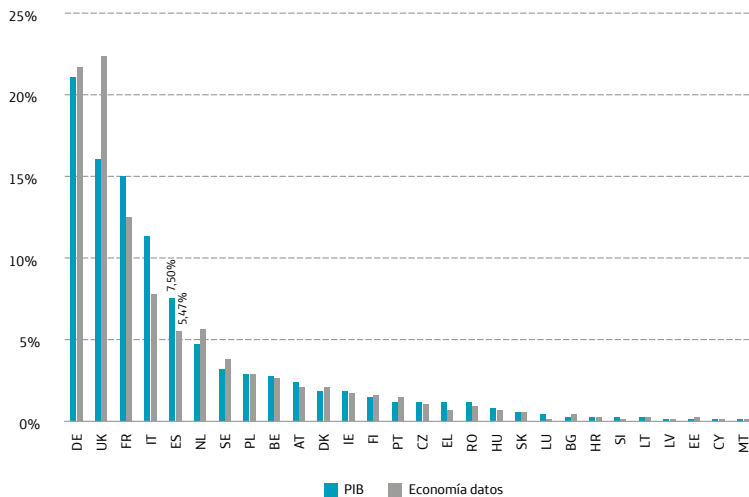
7. IDC y Open Evidence (2017): *European Data Market Smart 2013/0063*.

siderar el valor de la producción nacional, es decir, sumar el valor de los bienes y servicios producidos en el país y vendidos al exterior y descontar el valor de aquellos otros que han sido elaborados o prestados en otros territorios pero han sido comprados por empresas nacionales.

Según esta metodología, en el caso particular de España, se desprende que el tamaño del mercado de los datos es relativamente menor que el que le correspondería, a juzgar por el tamaño de su economía en el conjunto de la Unión Europea. El valor del mercado de los datos en España alcanzó en 2016 los 3.261 millones de euros, una cuarta parte de su peso en Alemania (12.925 millones de euros) o Reino Unido (13.313 millones de euros), que es la economía que lidera este *ranking* europeo. El relativo retraso de nuestra economía en este ámbito se refleja en que el PIB español representa el 7,5% de la economía de la UE-28, porcentaje que se reduce al 5,5% cuando se trata del mercado de los datos. Lo mismo les ocurre a otras economías europeas de gran tamaño, como Francia e Italia. En cambio, Reino Unido no solo lidera este *ranking* europeo, sino que, además, la relevancia del mercado de los datos es sustancialmente mayor que el peso de su economía en el conjunto de la UE-28.

Fig. 9.1

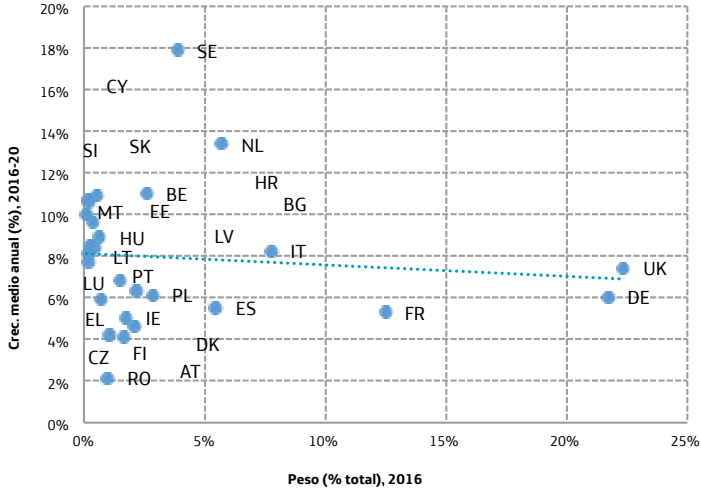
PIB y mercado de los datos en los países europeos (% total UE-28), 2016 (países ordenados de mayor a menor peso del PIB)



Fuente: Afi, IDC, Eurostat.

Fig. 9.2

Mercado de los datos en los países europeos (% total 2016 y crecimiento medio anual previsto para 2016-2020)



Fuente: Afi, IDC.

La Comisión Europea establece⁸ la necesidad de extender el valor del mercado de los datos a todos los agentes que participan en la economía, fomentando su uso, especialmente, en las pymes, los centros de investigación y las instituciones académicas. Aunque las perspectivas para el periodo 2016-2020 apuntan a un aumento medio anual del 5,5% del tamaño del mercado de los datos en España, será menor que el previsto para la media de la Unión Europea (7,5%) y los países líderes en esta materia en la actualidad, salvo en el caso de Francia. Por lo tanto, España podría perder la oportunidad de aprovechar la emergencia de la Economía de los Datos, y perder posiciones de liderazgo en el *ranking* europeo, si no impulsa decididamente la utilización y la comprensión de los mismos en nuestras empresas, centros de investigación, estudiantes y usuarios en general.

8. European Commission, European Strategy on the Data Value Chain (2016).

Bibliografía

- Arthur, Lisa. «What Is Big Data?». *Forbes*, 2013.
- Bitzstein, Joe, y Pfister Hanspeter. *Data Science Harvard Course*. 2015. <http://cs109.github.io/2015/>
- Bowden, Jason. «The Key Findings on Big Data Study». *Digital-Warriors Internet Marketing*, 2014.
- Buttarelli, Giovanni. «Data protection as a bulwark for digital democracy». *Keynote speech at the 6th International e-Democracy 2015 Conference on Citizen rights*. 2015.
- Carreño, Camila. «Big Data crecerá un 657% en los próximos 5 años en Latinoamérica». *Pulso Social*, 2014.
- CNMC. «Estudio sobre los nuevos modelos de prestación de servicios y la economía colaborativa». Madrid, 2016.
- ESMARTCITY. «Situación de las TIC en España». 2013.
- European Commission. «European Strategy on the Data Value Chain». DG Connect European Commission , 2016.
- Famian, Seth. *Working with big data*. 2016.
- Henke, Nicolaus, et al. *The Age of Analytics: competing in a data-driven World*. McKinsey Global Institute, 2016.
- Kroes, Natalie. «Data is the new gold». *Opening Remarks, Press Conference on Open Data Strategy*. Bruselas, 2011.
- Kurzweil, Raymond. *The singularity is near: When humans transcend biology*. Penguin, 2005.
- Manyika, James, Michael Chui, Peter Groves, Diana Farrell, Steve Van Kuiken, y Elizabeth Almasi Doshi. «Open data: Unlocking Innovation and Performance with Liquid Information». McKinsey Global Institute, 2013.

- Merriam-Webster. *Merriam-Webster's collegiate dictionary*. Merriam-Webster, 2004.
- Moore, Gordon Earl. «Cramming more components into integrated circuits». *Electronics* 38, n.º 8 (1965): 114.
- OECD. «Data-Driven Innovation for Growth and Well-being». Interim Synthesis Report, 2014.
- Poddar, Divya. «Big Data Analysis: Industries using it!». *UpX Academy*, 2016.
- Rotella, Perry. «Is Data The New Oil?». *Forbes*, 2012.
- SIIA. «Data-Driven Innovation. A guide for Policymakers: Understanding and Enabling the Economic and Social Value of Data». Public Policy Division, Software & Information Industry Association, 2013.
- The Economist. «The world's most valuable resource is no longer oil, but data». *The Economist*, 2017.
- Vickery, Graham. «Review of Recent PSI Re-Use Studies», 2011.
- White, Tom. *Hadoop, The Definitive Guide*. O'Reilly Media, 2015.

Economía de los Datos

Riqueza 4.0

Un elemento tan cotidiano y aparentemente inocuo como son los datos, en sus múltiples representaciones y formatos, está revolucionando la forma en la que hoy resolvemos los problemas, enfrentamos las preguntas que se nos presentan y abordamos tareas que hasta hace escasos años realizábamos de forma manual y analógica y con capacidades analíticas limitadas. Ejemplos ya observados en el mercado como la oferta personalizada de todo tipo de productos y servicios basada en el conocimiento que a través de los datos las empresas tienen de nosotros, consumidores; más incipientes como la individualización de los tratamientos médicos, o pronto de uso masivo como el control remoto de todos los dispositivos conectados de nuestro hogar (a través de Internet de las Cosas), nos ayudan a visibilizar algunas de las aplicaciones más inmediatas del uso de los datos.

La Economía de los Datos se está convirtiendo a pasos acelerados en un nuevo paradigma, y el dato en un nuevo insumo productivo, que acompaña los procesos de toma de decisiones en los ámbitos más relevantes de nuestra vida, tanto profesional como personal. Tomar conciencia de esta realidad que, por ser aparentemente invisible a nuestros ojos no significa que no exista, es hoy una tarea que todos y todas debemos abordar como ciudadanos que somos del siglo XXI. La información almacenada en el mundo (centros de datos) crece a un ritmo exponencial y alcanza dimensiones que eran inimaginables hace pocos años: si en 2013 alcanzó los 161 exabytes (EB) por año (desde 5 EB en 2003), en 2020 se prevé que el tráfico de datos supere los 15 zettabytes (ZB).

Estar preparados para un futuro que ya se vislumbra, materializado en una de sus múltiples facetas en la Economía de los Datos, es condición necesaria para aprovechar y optimizar las bondades que trae consigo la transformación digital por la que nos encontramos ya transitando, tránsito que la globalización nos exige acelerar para no quedarnos rezagados como sociedad en un contexto de creciente competencia, en su doble acepción del término.

Esta publicación editada por Fundación Telefónica, pretende acercar, animar a explorar y familiarizar al lector con los elementos básicos de uno de los principales vectores de crecimiento de la era digital en la que estamos inmersos, poblada de sucesivas innovaciones tecnológicas, muchas de ellas conocidas por el lector a través de la literatura de ciencia ficción. Innovaciones cuyo sentido o propósito es mejorar nuestro bienestar tanto por la vía de las ganancias de eficiencia como por la expansión de las fronteras del conocimiento aplicado.