

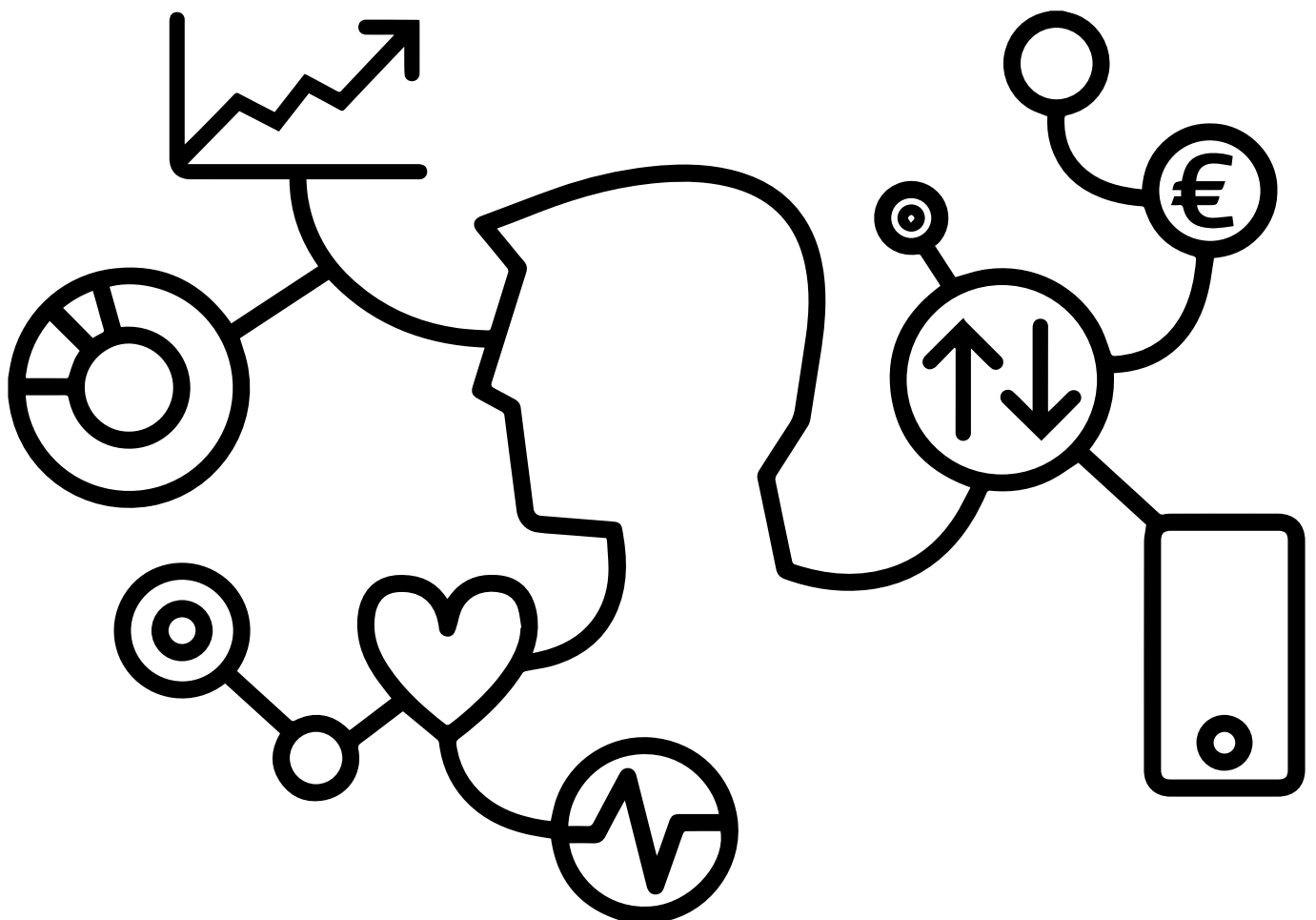


- > **Nuestra producción de datos ha crecido exponencialmente en pocos años.**
Por qué son necesarias herramientas que analicen y procesen la información relevante.
- > **Las implicaciones del Big Data para cada uno de nosotros como individuos y nuestras relaciones con el entorno.**
- > **Cómo mejorar nuestro rendimiento humano con el procesamiento de los grandes volúmenes de datos.**

Big data

El poder de los datos

"La tecnología big data ya está aquí, ahora hay que aprender a usarla."





Fundación
Innovación
Bankinter

Big data

El poder de los datos

Agradecimientos

Nuestro agradecimiento a **Esther Paniagua**, periodista y autora de este informe. Su trabajo ha sido decisivo para poder plasmar las conclusiones de la esta tendencia del Future Trends Forum.

Nuestro agradecimiento a todos los miembros del Future Trends Forum (FTF) que han hecho posible el éxito de nuestra última reunión, especialmente a aquéllos que han participado activamente en la realización de esta producción:

Por su inestimable colaboración en la elaboración de esta publicación:

Chris Meyer
Lita Sands
Vicki Sayfert-Margolis
Stefan Bungart
J.C. Herz

En la organización y metodología de la reunión del Future Trends Forum:

Chris Meyer

Y por último, agradecer a las personas del equipo, por su compromiso y buen hacer en el desarrollo del contenido de esta publicación:

Fundación Innovación Bankinter
Sergio Martínez-Cava
Marce Cancho
María Teresa Jiménez
Lara García de Vinuesa
Dorsey Lockhart
Pablo Lancry



Garrick Jones
Clemens Hackl



Raquel Durán

Las opiniones expresadas en este informe son del autor y no reflejan la opinión de los expertos que participaron en la reunión del Future Trends Forum.

Ponentes y asistentes

Ana María Arboleda

Directora General de Atlantis Healthcare Spain

Miguel Arias

COO en CartoDB

Puneet Batra

Cofundador de LevelTrigger

Jessica Bland

Investigadora de Tecnologías del Futuro de Nesta

Carla Brodley

Decana de la College of Computer and Information Science de la Northeastern University

Stefan Bungart

Líder de GE Software Europe en GE Global Research

Antonio Damasio

Profesor y Director del Brain and Creativity Institute en la Universidad del Sur de California (USC) y Premio Príncipe de Asturias de Investigación Científica y Técnica

Hanna Damasio

Directora del Dana and David Dornsife Cognitive Neuroscience Imaging Center de la Universidad del Sur de California (USC) y Premio Príncipe de Asturias de Investigación Científica y Técnica

Maarten den Braber

Co-fundador de SHIFT

Julie Freeman

Artista y líder de Datos del Culture Art Programme en el Open Data Institute (ODI)

J.C. Herz

Diseñadora y consultora estratégica con Big Data

Eduardo Hruschka

Jefe Científico de Datos en Big Data Brazil

Richard Kivel

Emprendedor e inversor en Tecnología médica y biofarmacia y Patrono de Fundación Innovación Bankinter

Joseph Kvedar

Director del Center for Connected Health

Gregory La Blanc

Investigador Haas Economic Analysis en la Universidad de Berkeley

Philip Lader

Presidente no ejecutivo de WPP Group y Patrono de Fundación Innovación Bankinter

Julia Li

Fundadora y CEO de HCD Global

Juan Carlos López

Líder de Relaciones Académicas y colaboración de Roche Innovation Center

Chris Meyer

CEO de Nerve LLC y Patrono de Fundación Innovación Bankinter

Emilio Méndez

Director del Center for Functional Nanomaterials del U.S. Department of Energy's, Brookhaven National Laboratory, Premio Príncipe de Asturias 1998 y Patrono de Fundación Innovación Bankinter

Esteban Moro

Investigador en la Universidad Carlos III de Madrid

Iyad Rahwan

Profesor de Tecnologías de la Información en el Masdar Institute

Pablo Rodríguez

Investigador y Director de Innovación en Telefónica

Lita Sands

Líder global de Transformación Digital de Novartis

Vicki Seyfert-Margolis

Fundadora y CEO de My Own Med, Inc.

Eden Shochat

Fundador de Aleph y Patrono de Fundación Innovación Bankinter

Taryn Sullivan

CEO de Efficiency Exchange

Steve Trachtenberg

Presidente Emérito de la George Washington University y Patrono de Fundación Innovación Bankinter

Wilfried Vanhonacker

Profesor en Olayan School of Business, AUB y Patrono de Fundación Innovación Bankinter

Andreu Veà

Presidente de la Internet Society (ISOC-ES)

Muchas gracias,
Fundación Innovación Bankinter

El poder de los datos

➔ **Nuestra capacidad** para producir información ha avanzado a la velocidad de la luz con respecto a hace tan solo unos años. La ingente cantidad de datos disponibles ha hecho necesario el desarrollo de herramientas que los analicen y procesen para identificar lo relevante. Estas tecnologías ya existen pero apenas hemos empezado a exprimir sus posibilidades.

"Si se pusieran en fila todos los nuevos libros publicados, nos deberíamos desplazar a 150 kilómetros por hora para mantenernos al frente de la hilera", escribía Stephen Hawking en *El universo en una cáscara de nuez* hace ya trece años. "Si continuara el crecimiento exponencial, se publicarían diez artículos por segundo en mi especialidad de física teórica, y no tendría tiempo de leerlos", proseguía el aclamado físico.

En efecto, eso ha sido exactamente lo que ha pasado. Según el profesor de bioinformática de la Escuela de Salud Pública de Harvard Winston Hide, **en los últimos cinco años se ha generado más información científica que en toda la historia de la**

humanidad. El volumen de los datos existentes es de tal magnitud que, si ocupara un espacio físico, Hide asegura que superaría el tamaño de una galaxia.

Esta aceleración de la velocidad a la que generamos contenido ha propiciado la emergencia de una nueva ciencia: el *big data*, que en español se podría traducir como "datos masivos". Su significado parece claro, sin embargo, este concepto no es suficiente per se para explicar el fenómeno, que trasciende a una simple idea de cantidad. No estamos hablando solo de un montón de datos, sino de la generación de valor a partir de su procesamiento y análisis. Esa es la verdadera **revolución del big data** y el motivo que ha llevado a una treintena de expertos de todo el mundo

a reunirse en la XIII edición del *Future Trends Forum* organizado por la Fundación Innovación Bankinter, que inspira esta publicación.

Para llegar hasta aquí se han desarrollado numerosos dispositivos y aplicaciones que permiten medir, estructurar, procesar y analizar grandes volúmenes de datos gracias a nuevas infraestructuras tecnológicas diseñadas para soportar *petabytes* de información. Constantemente aparecen nuevas y refinadas tecnologías que están permitiendo mejorar la eficiencia de las organizaciones, precisar los gustos y necesidades de clientes y consumidores, conectar a Gobiernos y ciudadanos, reforzar la seguridad y la inteligencia de las ciudades, realizar predicciones de todo tipo y generar **patrones de comportamiento** útiles para que las empresas conozcan mejor a su público y para concienciar a las personas sobre sus propios hábitos, incluso dando recomendaciones de cómo cambiarlos para mejorar aspectos de su vida como la salud.

Todo esto es solo el comienzo, ya que apenas estamos empezando a aprovechar el potencial que ofrece el análisis de datos masivos. El *big data* ha abierto la veda a una sociedad más informada, más eficiente, capaz de realizar proezas hasta hace poco reservadas al ámbito de la ciencia ficción.

Tal vez Hawking no tenga tiempo de leer todos los artículos científicos que se generen, pero tampoco le hará falta. Alguna tecnología *big data* lo hará por él, y solo tendrá que preocuparse de interpretar los resultados. De hecho, no tendrá que esperar mucho: trece años después de sus vaticinios, esto ya forma parte del presente. La **toma consciente de decisiones** basadas en datos es hoy una realidad, y quien no la aplique se quedará atrás.

El reto ahora es encontrar la aguja en el pajar, identificar **lo relevante** y aplicarlo involucrando a todos los agentes implicados (empresas, organizaciones, Gobiernos, profesionales y ciudadanos), superando las barreras legales, estructurales y humanas que presenta. Nadie dijo que fuera fácil, pero merece la pena intentarlo.

En los últimos cinco años se ha generado más información científica que en toda la historia de la humanidad.

Analizar grandes volúmenes de datos gracias a nuevas infraestructuras tecnológicas diseñadas para soportar *petabytes* de información.

Identificar lo relevante y aplicarlo involucrando a todos los agentes implicados.

Las seis W del big data para empresas

➔ **¿Para qué sirve** el *big data* en el entorno empresarial? ¿Qué puede hacer por mi compañía? ¿Cómo puedo aprovechar al máximo su potencial?

Toda noticia debe dar respuesta a, al menos, seis preguntas. Cinco de ellas empiezan en inglés por uve doble (*who* o quién, *what* o qué, *where* o dónde, *when* o cuándo y *why* o por qué) y la otra (how o cómo) también contiene esta letra. De ahí que se llamen las seis W del periodismo, aunque no son exclusivas de esta disciplina. Por eso hemos tomado prestada esta máxima para explicar la visión de ciencia de datos desde el punto de vista del negocio, respondiendo a esas seis cuestiones:

El '**quién**' está claro: el **big data**. Una industria que en 2010 valía 3,2 billones de dólares y que este 2015 alcanzará aproximadamente los 17 billones, [según IBM](#). Es un negocio a veces controvertido, sobre todo en términos de reputación para las empresas que hurgan en los datos que generamos en nuestra vida personal *online* y comercian con ellos. Pero también es rentable: empiezan a emerger en este sector empresas y *startups* como Acxiom o Splunk que no bajan de los [300 millones de dólares en ingresos anuales](#) y del [billón de dólares](#), respectivamente.

Al '**qué**' responde su utilidad: **cuantificar, predecir y cambiar**. Cuantificar para acceder a nuevos conocimientos al observar los resultados del análisis de volúmenes masivos de datos, a los cuales de otra

manera no habríamos podido acceder; para darnos cuenta de detalles importantes, de fugas que se nos escapaban o de patrones de comportamiento de las personas asociadas a la organización (empleados, clientes, consumidores...). Y en base a ello podremos predecir, por ejemplo, decisiones de compra o incluso de voto, o modificar procesos ineficientes.

¿Dónde? Dentro de la empresa podemos aplicarlo **de manera integral y transversal en los diferentes departamentos** (recursos humanos, operaciones, finanzas, marketing...) y en respuesta a diferentes objetivos: eficiencia de procesos y rendimiento, liderazgo y formación de talento, desarrollo de negocio, investigación e innovación en productos y servicios...

¿Cómo? Innovando. Con una visión amplia y aperturista que requiere un cambio de mentalidad corporativa. Integrando el *big data* en el ADN de la organización. Imaginado (y poniendo en marcha) nuevas formas de aproximarse al consumidor, de generar valor en el entorno al tiempo que se desarrolla un producto o que se presta un servicio -o de forma paralela- o creando nuevas líneas de negocio dirigidas específicamente a este objetivo; abriendo los problemas de la organización y

compartiendo sus retos para buscar soluciones colectivas; pasando de la tradicional innovación iterativa a una innovación disruptiva.

El '**por qué**' y el '**para qué**' están claros: para **optimizar y para transformar**. Los datos procedentes de internet, de medios de comunicación sociales y de nuestros teléfonos y dispositivos permiten medir el rendimiento con mayor precisión y usar esos conocimientos para predecir el rendimiento futuro. Estas mediciones sirven para hacer más eficientes los procesos de negocio pero a menudo piden ir más allá y transformar la estructura de la organización para hacerla más ágil y adaptable a las demandas del mercado.

Sin embargo, la ciencia del *big data* no puede hacer todo esto por sí sola. A las 6 'w' les falta el **factor humano**, que debe combinarse con el tecnológico.

Falta el factor humano, que debe combinarse con el tecnológico.

Por una parte tenemos herramientas cada vez más avanzadas de gestión financiera, análisis predictivo y causal e inteligencia empresarial que podemos integrar en una sola una plataforma. Pero estas aplicaciones no servirán de nada si el cambio no se da de forma paulatina y teniendo como punto de partida la escucha activa a los empleados, en un proceso de medición constante de sus necesidades y de refuerzo continuo de los valores y motivos de la transformación que se está llevando a cabo.

Unidos por los datos

➔ El **big data** nos conecta con las personas a nuestro alrededor y potencia la interacción a diferentes niveles -del social al empresarial- y en diferentes ámbitos -de la salud a la seguridad.

De las empresas pasamos a las personas. Hablamos de las **implicaciones del big data para cada uno de nosotros como individuos y en nuestras relaciones** con el entorno: entre gobernantes y ciudadanos, entre médicos y pacientes, entre directivos y empleados, entre jugadores y entrenadores o de las fuerzas de seguridad contra los criminales.

La tecnología está acelerando la capacidad de escuchar los problemas. Eso tiene un gran interés en el **funcionamiento de las ciudades**. Las emergentes plataformas de participación puestas en marcha por numerosos gobiernos en todo el mundo -y sea a nivel local, regional o nacional- permiten expresarse a los ciudadanos acerca de sus necesidades con



Implicaciones del *big data* para cada uno de nosotros como individuos y en nuestras relaciones.

respecto al lugar que habitan y permiten también notificar incidencias. Como contrapartida, se generan expectativas de respuesta rápida por parte de los poderes públicos.

El problema no es que no se pueda proporcionar esa respuesta sino que a lo mejor no es lo prioritario. La presión ciudadana puede llevar a tomar decisiones inadecuadas y dejar de lado la implementación de soluciones a problemas de largo alcance que requieren una mirada más pausada y una actuación más concienzuda y sostenida en el tiempo.

El sistema sanitario es otro gran vertebrador que está notando la llegada del *big data*. De momento lo hace como un espectador más, como si la cosa no fuera con él. Pero mientras actúa como convidado de piedra todo a su alrededor se mueve y **cambian las relaciones entre pacientes, cuidadores y profesionales sanitarios** dentro y fuera del sistema, en las comunidades de proximidad (físicas o cibernéticas) y en la unidad familiar. También lo hace el autocuidado a nivel individual, con la toma de conciencia de nuestros hábitos y de cómo podríamos mejorarlos a partir de aplicaciones y dispositivos de medición de actividad y de marcadores de salud. Aunque no quiera darse cuenta, el sistema sanitario ya se está reinventando.

Lo están haciendo también las corporaciones, donde el *big data* ha llegado como el acicate que

faltaba para dar el paso hacia la renovación de unas estructuras jerárquicas añejas que no satisfacen las necesidades actuales de los empleados. Se ha iniciado una **transición hacia una nueva forma de organización -y de relaciones** entre las personas que trabajan en una organización- donde la transparencia, a través de una política de datos abiertos, es un valor clave.

Los empleados exigen mayor autonomía en la toma de decisiones y el *big data* les proporciona un cierto empoderamiento que a su vez desemboca en una mayor toma de conciencia sobre sus responsabilidades. Al mismo tiempo se redefine el rol de los poderes directivos, que pasan de ser jefes a ser líderes, de ser autoritarios a ser facilitadores, de fiscalizar a potenciar el talento.

Otro papel que está cambiando es el de los entrenadores deportivos. La ciencia computacional puede ayudarles desde su concepción del juego no como un conjunto de individuos sino como una red cuyo desempeño es óptimo cuanto mejor es la conexión (los pases) de los puntos que la componen (los jugadores). Por tanto, se podrían definir nuevas estrategias para el campo de juego combinando el análisis de datos y la teoría de redes. De hecho, ya se hace. El obstáculo, más que tecnológico, es humano: requiere una actitud abierta hacia los datos y una **democratización de la relación entre jugadores y entrenadores**. La labor de estos últimos será ahora la de "aplicador" de los resultados obtenidos del análisis de datos, trasladándolos a una estrategia que transmitirá a los jugadores. Y gran parte del trabajo previo corresponderá a una nueva figura: la del científico de datos, que se encargará de su recopilación y limpieza.

Esta figura escasea porque se trata de una profesión emergente y porque la demanda aún no se ajusta a las necesidades reales. Queda patente, por ejemplo, en el caso de las fuerzas de seguridad. Entre los 49.000 policías de Nueva York apenas hay 880 entrenados en datos. Eso les deja en una situación poco favorable a la hora de utilizar las últimas herramientas a su alcance para prevenir y detectar

delitos y crímenes (tanto *online* como en la calle) y para encontrar a los culpables. Aunque cuenten con la más avanzada tecnología, la ausencia de personal capacitado para usarla lo hará inútil.

Esta escasez de recursos contrasta con la abundancia económica en la que se bañan las bandas criminales, lo que les permite acceder a las mismas herramientas

que la policía. A esto se añade el hecho de que no se atienen a barreras legales de privacidad, protección de datos o de intercambio de información. Es decir, pueden usar el *big data* sin límites para desarrollar superorganizaciones criminales inteligentes y seguir aumentando su red y su cadena de suministro. Por tanto, **la relación entre policías y criminales se da en una situación de injusta desventaja.**

Asignaturas pendientes

➔ **Pulir los datos**, comprenderlos y aplicarlos garantizando la privacidad son los tres grandes retos del *big data*.

Al *big data* le han quedado tres para septiembre y corre peligro de suspenderlas si no se actúa en consecuencia. La primera tiene que ver con la limpieza de los datos basura. La segunda, con la comprensión lectora y su aplicación práctica. Y la tercera con la ética y la responsabilidad. Para que estas materias no pasen de reto a obstáculo y nos hagan repetir de curso, hay que hacer una serie de deberes:

Contextualizar y limpiar. Es lo primero que hay que hacer con los datos para que sus aportes sean de calidad y no meros declarativos inútiles. Los algoritmos se tienen que nutrir de la interacción humana porque no significa lo mismo tener una frecuencia cardíaca elevada mientras se practica ejercicio físico que estando en el sofá. Aún así, siempre se colarán datos inservibles, desestructurados y erróneos que habrá que detectar y eliminar. Este

trabajo será más sencillo cuanto mejor se hayan filtrado y seleccionado anteriormente, sin abusar de la factibilidad técnica de recopilar sin límite información que en realidad no necesitamos para nada.

Entender y aplicar. Es la segunda tarea. Una buena comprensión lectora de los datos es clave para transmitirlos adecuadamente. Para esto son necesarios científicos de datos y analistas cualificados que trasladen los mensajes importantes, y para ello pueden valerse de herramientas de visualización que ya están disponibles y que ayudan a que los profesionales de los negocios entiendan el *big data*. El siguiente paso es la aplicación de ese conocimiento, saber qué hacer con esos datos. No es tan fácil como parece. Profesionales como los sanitarios se enfrentan diariamente a dudas acerca del uso que deben dar a los datos de los pacientes

(mejorar el abordaje, aprender sobre lo que funciona y lo que no del sistema...).

Equilibrar. Por otra parte, a la hora de recopilar estos datos médicos surge otra incógnita: ¿dónde parar? Esto tiene que ver con la tercera materia pendiente: hacer de contrapeso entre propiedad, privacidad y seguridad en una sensible balanza donde también entra en juego el concepto de propiedad y su delimitación. ¿A quién pertenecen los datos que gobiernos, fuerzas de seguridad y otras organizaciones recogen de las personas? ¿Dónde

empieza el derecho de unos a usarlos y dónde acaba el derecho de otros a la protección de sus propia información? Por una parte, está la privacidad de las personas y la seguridad de sus datos, pero esta se contraponen a su seguridad física cuando interfiere en una acción policial que podría beneficiarse de disponer de dichos datos.

Claramente hay un conflicto de intereses que es preciso resolver si queremos beneficiarnos del *big data*. No hay elección: o hacer los deberes o catear las asignaturas pendientes.

La tierra prometida

➔ **Cualquier tiempo pasado** no fue mejor. Lo será el futuro si las aspiraciones de los grandes datos no se quedan en *wanna be*.

Hacer frente a los retos y superar las barreras y obstáculos del *big data* puede llevarnos a un lugar de ensueño. Un mundo totalmente conectado donde impere la transparencia; donde todos los datos sean abiertos y las herramientas libres y colaborativas; donde empresas, organizaciones e instituciones, gobiernos e individuos luchen de manera conjunta contra la pobreza, los problemas sanitarios, alimenticios y medioambientales; donde el trabajo colaborativo y la innovación abierta conduzcan a la creación y materialización de tecnologías de ciencia ficción.

Esa época dorada podría llegar si se superan los **obstáculos legales, humanos, de mercado, de formatos, de acceso y de habilidades** que hoy lo impiden. La historia del futuro empieza a escribirse con una gran alianza mundial de compartición de

datos contra los grandes problemas mundiales, facilitada por una normativa sobre privacidad a nivel global que armonizará el acceso a los datos. El control de su cumplimiento se realizará mediante algoritmos de aprendizaje automático.

A medio plazo se liberarán las herramientas de manejo de datos y años después adquirirán la forma de una aplicación móvil con apariencia de juego que hasta los niños dominarán. Al mismo tiempo, se abrirán los datos bajo un único estándar internacional y se homogeneizarán formatos como por ejemplo los de los historiales médicos. Pero, en medio de todo esto, ocurrirá una catástrofe: un ciberataque pirata de una superorganización criminal destruirá la red informática mundial y con ella internet tal y como lo conocemos.

Un mundo totalmente conectado donde impere la transparencia; donde todos los datos sean abiertos y las herramientas libres y colaborativas.

Esto obligará a replantear las fórmulas de protección y privacidad y cómo combatir el fraude y el robo de identidad. Volveremos una posición conservadora y de resistencia social, que se reducirá a medida que avancen la transparencia y los sistemas que garanticen la privacidad de datos. También habrá cada vez más gente interesada en formarse como científico o como analista de datos, y **se crearán nuevas disciplinas educativas** dentro y fuera del sistema de educación oficial, *offline* y *online*, y ningún universitario terminará su grado sin haber estudiado la ciencia del *big data*. El empleo estará asegurado al menos unos años, hasta que el exceso de oferta sature el mercado.

Las ciudades podrán autogestionarse a nivel operativo a través de algoritmos y las empresas contarán con un mercado donde podrán comprar y vender datos. Y a medida que avance la tecnología de sensores y los datos sean de más alta calidad se podrán materializar esperados productos como los coches conectados autoconducidos, que cambiarían la visión del automóvil como un bien para convertirse en un servicio. También habrá ordenadores cuánticos al servicio de los datos. Y el problema del volumen no existirá porque todos los datos del mundo cabrán en un disco duro del tamaño de una cuchara, como cabe el universo de Hawking en una cáscara de nuez.

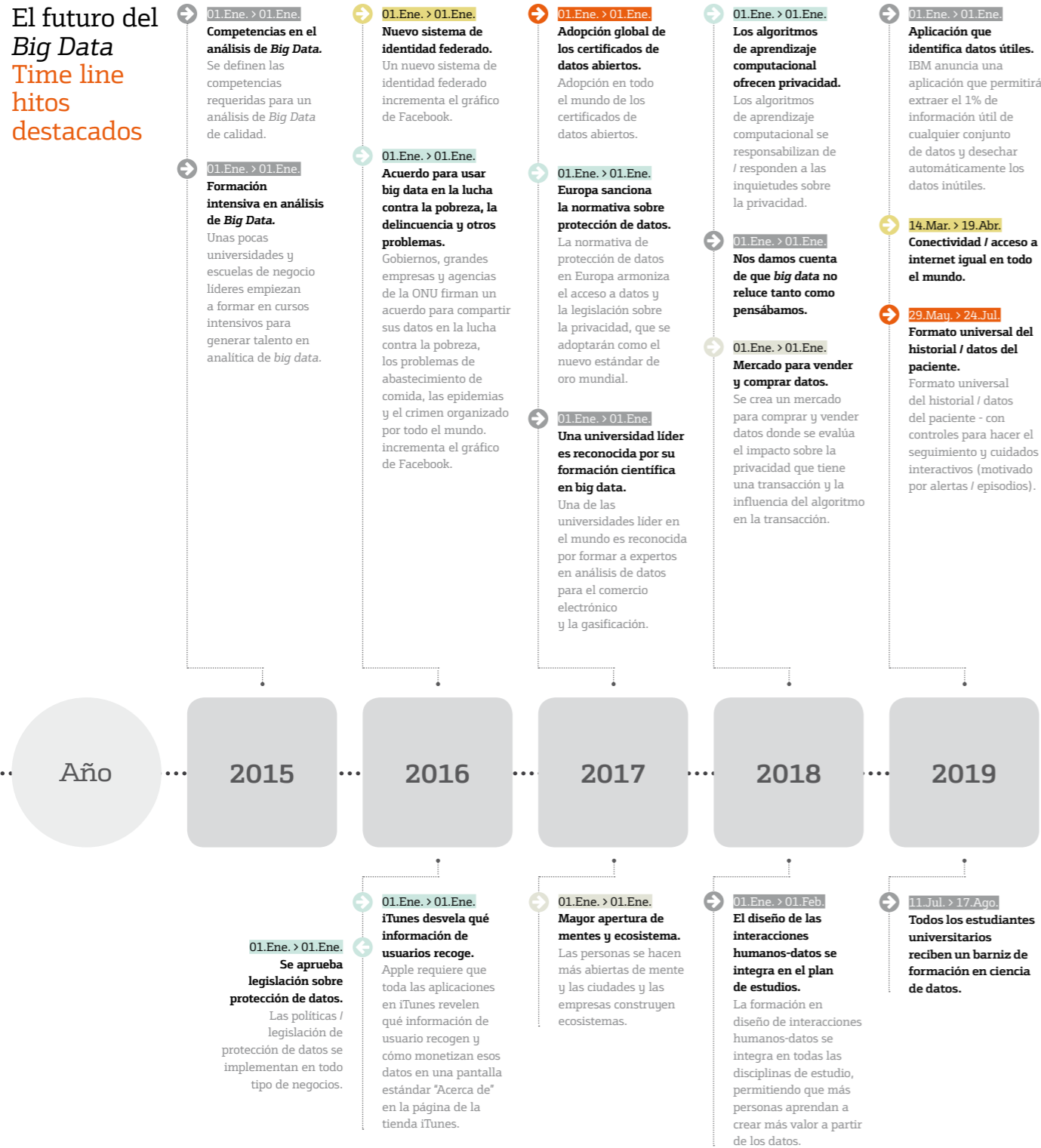
Es el futuro que nos espera si los expertos que hacen estas predicciones no van muy desencaminados. Pero ahora es tiempo de volver a la realidad de 2015 y empezar a trabajar para disfrutar, en unos pocos años, de la época dorada de los grandes volúmenes de datos.

El futuro del Big Data

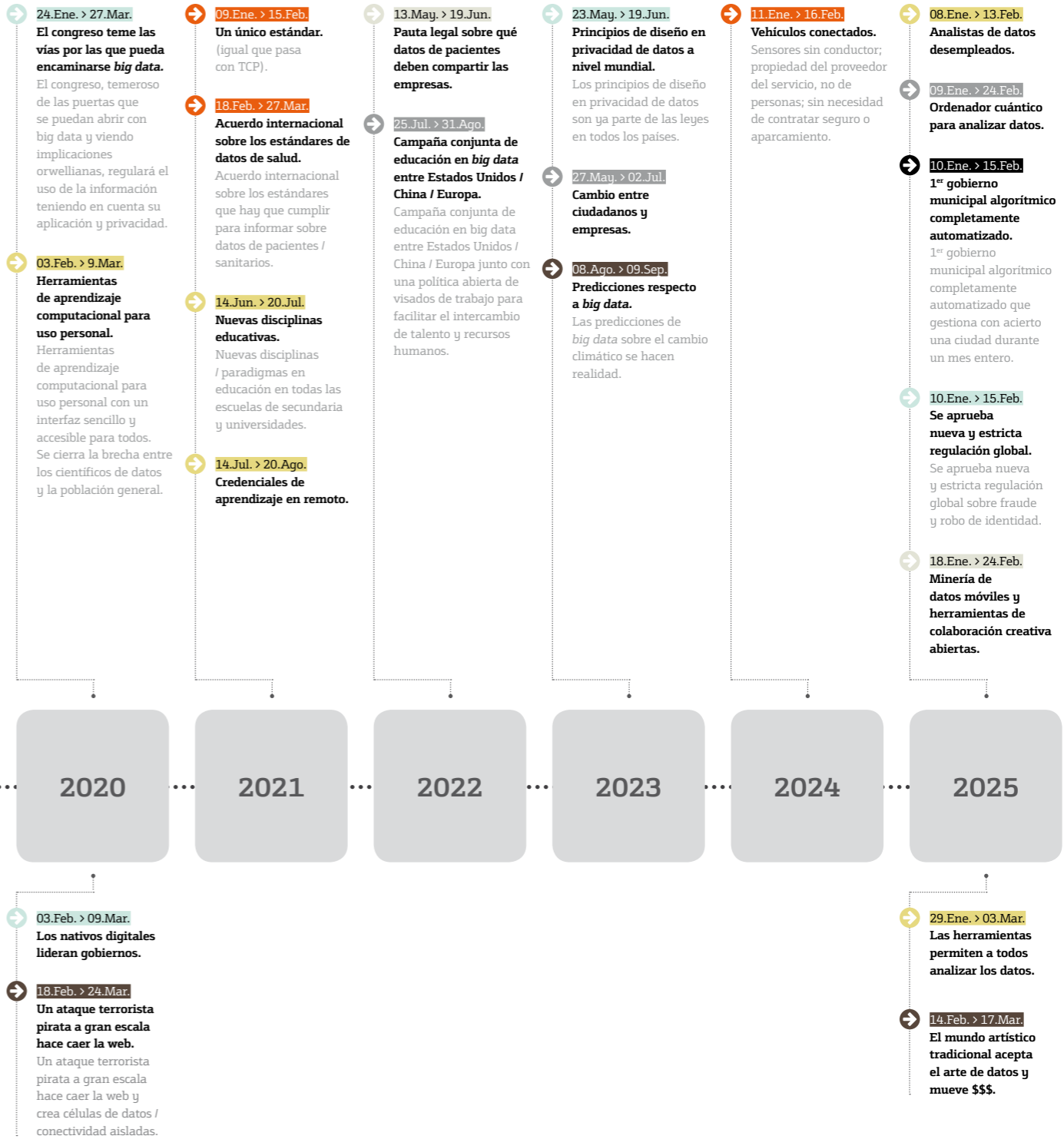


El futuro del Big Data

Time line hitos destacados



- ➔ Hitos en mercados por desarrollar
- ➔ Hitos en formatos, calidad y estándares de datos
- ➔ Hitos en relación con el factor humano y la resistencia que se pueda oponer
- ➔ Hitos en legislación y marco legal



- ➔ Hitos en aptitudes y recursos
- ➔ Hitos respecto a los compartimentos estanco (Acceso, API)
- ➔ Otros hitos



