

Aspectos legales del *Big Data*

Carlos Pérez Sanz
Socio. Ecija Abogados

INTRODUCCIÓN

Concepto de *Big Data*

Son muchas las definiciones que podemos encontrar sobre el concepto de *Big Data*, pero todas ellas tienen en común describir a este fenómeno tecnológico como el conjunto de recursos que permiten la gestión y análisis de cantidades ingentes y masivas de datos, con un alcance y dimensiones en constante crecimiento y sin parangón en la historia de la tecnología.

En la definición del *Big Data*, también se acepta comúnmente su caracterización a través de las denominadas “las 3 uves”: volumen, variedad y velocidad.

De entre los recursos que integran el fenómeno *Big Data*, los siguientes son imprescindibles:

- **Recursos humanos**, cuya aportación creativa es muy variada. Esta requiere, en primer lugar, personas con amplios conocimientos técnicos, que pueden suponer, desde las tareas más elevadas, las del denominado “*data scientist*” o científico de datos, cuya labor consiste en la concepción y desarrollo de algoritmos que aportan valor añadido diferencial al proyecto de *Big Data* en cuestión, hasta la mera tarea de depuración de datos, pasando por la definición de objetivos y finalidades del proyecto o por la elección de las mejores técnicas de combinación o de anonimización de datos. En segundo lugar, el proceso de tratamiento de datos y la implemen-

tación de un proyecto *Big Data* requiere personas con un gran conocimiento del negocio y del sector en el que el proyecto vaya a operar.

- **Infraestructura tecnológica**, integrada por arquitectura hardware y software con dimensión y potencia suficientes y que, actualmente, suelen incluir soluciones de computación en la nube, programas de ordenador de gestión, selección y presentación de los datos, así como los algoritmos que aportan la lógica diferencial del proyecto.
- **Fuentes de datos**, cuya variedad, accesibilidad y capacidad de combinación e interrelación es, en el estadio tecnológico actual, de dimensiones enormemente superiores a las disponibles en estadios tecnológicos anteriores. Dichas fuentes incluyen los sistemas de captación de información, las bases de datos y repertorios históricos de datos propios del impulsor del proyecto *Big Data*, así como fuentes de datos ajenas a dicho impulsor (Internet, redes sociales, información puesta a disposición general por las administraciones públicas, etc...).

Elemento diferencial

El elemento diferenciador del fenómeno de *Big Data* es su **capacidad predictiva**. Hasta la aparición y consolidación del fenómeno *Big Data*, y por las limitaciones propias del estado de la técnica del momento, los proyectos relativos a inteligencia de datos se limitaban, en gran medida, a analizar la información histórica propia disponible, dándole orden y sentido, con el fin de poder diseñar de modo más eficaz los propios procesos productivos o de toma de decisiones.

Por contra, los proyectos *Big Data*, por el enorme desarrollo, potencia y capacidad combinativa de los diversos recursos involucrados, permiten a su impulsor anticiparse a acontecimientos futuros y, en algunos casos, predecirlos con poco margen de error. Siguiendo con el ejemplo anterior, un proyecto *Big Data* ayudaría a la organización a analizar el carro de la compra de sus

La privacidad es uno de los aspectos fundamentales que deben considerarse a la hora de analizar los proyectos Big Data desde un punto de vista jurídico

clientes y predecir si uno de ellos comenzará a aumentar su consumo de cerveza y pizza en función de su estilo de vida, o si por el contrario sus compras anteriores indican que vive en pareja y podrá renovar pequeños electrodomésticos como la batidora. Este conocimiento tan detallado permite a nuestra compañía, por ejemplo, enviar promociones personalizadas antes que los competidores, adelantándose así al deseo de compra de los clientes y aumentando las ventas anuales.

Objetivos principales

Las finalidades perseguidas por un proyecto *Big Data* pueden ser muy variadas. Ahora bien, la práctica totalidad de dichos proyectos pueden incluirse en alguna de las siguientes categorías, en función del objetivo principal perseguido:

- **Inteligencia comercial**, que engloba a todos aquellos proyectos *Big Data* que persiguen un conocimiento profundo y predictivo de los clientes que interactúan con el impulsor del proyecto. Estos proyectos buscan mejorar la experiencia del cliente en su relación con el impulsor, prevenir y evitar que el cliente corte la relación con el impulsor, generar patrones de comportamiento y segmentaciones avanzadas de clientes o la oferta personalizada de productos, servicios y precios.
- **Fraude y riesgo**: consistentes en la previsión y predicción de posibles actividades fraudulentas que redunden en perjuicio del impulsor del proyecto, desde las de sus propios clientes y proveedores, hasta de otros elementos externos (ciberseguridad o prevención de fallos de sistema).
- **Eficiencia operativa**: encaminados a optimizar los procesos productivos o de toma de decisiones, incluyendo proyectos asociados al Internet de las Cosas, eficiencia en la gestión logística y de flotas, o de optimización en la gestión de reclamaciones y devoluciones.
- **Monetización**: por último, están los proyectos *Big Data* que persiguen un beneficio económico directo, a través de la generación de nuevos negocios en entorno digital, apps, empresas *fin-tech*, intermediación en *e-commerce*, etc...

Sobre las anteriores premisas conceptuales, se analizarán los aspectos más importantes del fenómeno *Big Data* desde el punto de vista jurídico.

Existen muchas y muy variadas metodologías de anonimización, y es muy posible que sea necesario combinar varias de ellas para asegurar la completa disociación en un proyecto Big Data

ASPECTOS JURÍDICOS PRINCIPALES

Privacidad y protección de datos personales

Es indiscutible que la privacidad es uno de los aspectos fundamentales que deben considerarse a la hora de analizar los proyectos *Big Data* desde un punto de vista jurídico, se analizan a continuación los aspectos más importantes de este ámbito normativo y su impacto en proyectos de esta naturaleza.

Finalidades legítimas

Uno de los principios orientadores de la normativa de privacidad es el de adecuación a las finalidades de tratamiento, según el cual los datos personales no pueden tratarse para finalidades incompatibles para aquellas que fueron informadas a los interesados en el momento de recabar sus datos.

Buena parte de los proyectos *Big Data* corre el riesgo de incumplir dicha salvaguarda para la privacidad personal, bien porque la combinación de los datos ya disponibles con nuevas fuentes de datos adicionales permite generar perfiles personales sin que exista **consentimiento previo** para ello, bien porque los objetivos del proyecto conducen a usos de datos que no eran razonablemente previsibles para los interesados en el momento inicial de obtención de sus datos. Estas situaciones de transgresión a la privacidad conllevan el correspondiente riesgo sancionador, incrementado de forma muy sustancial con el nuevo Reglamento Europeo de Protección de Datos que, si bien no será de aplicación hasta el 25 de mayo de 2018, prevé sanciones administrativas muy superiores a las actuales, que pueden alcanzar los 20 millones de euros o el 4% de la facturación total anual de la compañía.

Para superar estas situaciones de riesgo, las autoridades europeas en materia de protección de datos recomiendan someter el proyecto *Big Data* a un test de incompatibilidad, en el cual se sometan a prueba todos los elementos en juego:

finalidades informadas, contexto de obtención de datos, finalidades ulteriores del propio proyecto *Big Data*, tipología de datos involucrada y su impacto sobre los interesados conforme a las finalidades ulteriores del proyecto *Big Data*, etc. Dicho test se superará con éxito si se cumple alguna de las siguientes condiciones:

- Que las finalidades del tratamiento de datos del proyecto *Big Data* se ajusten a lo informado a los interesados en el momento inicial de recabar sus datos; o bien
- Que las finalidades del tratamiento de datos del proyecto *Big Data* sean razonablemente previsibles para los interesados, aun no habiendo sido explícitamente informados en el momento de obtener sus datos; o bien
- El tratamiento de datos resultante del proyecto *Big Data* está justificado por otras causas de legitimación previstas en la normativa de privacidad (como son, por ejemplo, el interés legítimo del responsable del tratamiento, el cumplimiento de obligaciones legales, contractuales, o en atención al interés vital de los interesados).

En caso de superarse con éxito este test de incompatibilidad, el proyecto de *Big Data* podrá considerarse conforme a la normativa de protección de datos, sin perjuicio del cumplimiento de otras obligaciones previstas por dicha normativa (como son la aplicación de medidas de seguridad, o la realización de análisis previos de impacto y su aprobación por las autoridades nacionales de protección de datos si el proyecto implica la generación de perfiles personales).

Ahora bien, si el test de incompatibilidad es negativo, el proyecto deberá sujetarse a información y consentimiento previos de los interesados involucrados, o bien someterse a procesos de anonimización y disociación de los datos.

Anonimización de datos

No todos los proyectos *Big Data* involucran tratamientos de datos de carácter personal, vinculados a personas físicas identificadas o identificables. Algunos no tienen como objetivo en absoluto a personas ni individuos, en otros la identidad de los individuos involucrados es completamente irrelevante, y en otros se hace necesario prescindir de la identidad de la persona por no haberse superado con éxito el test de incompatibilidad antes descrito.

Para las dos últimas situaciones, la aplicación de **técnicas de anonimización y disociación de datos son la mejor solución** para asegurar que el proyecto *Big Data* queda fuera del ámbito de aplicación de la normativa de protección de datos y de cualquier riesgo sancionador por incumplimiento de dicha normativa.

Ahora bien, para asegurar lo anterior, es necesario que la técnica de disociación aplicada garantice que se producen los siguientes efectos:

- **Aislamiento**, de forma que los registros de información que identifican a las personas quedan totalmente aislados y separados del resto de información objeto del proyecto *Big Data*;
- **No vinculación**, de forma que, tras el aislamiento, sea imposible volver vincular la información global del proyecto con los registros que aportan la identificabilidad de la persona; y
- **No inferencia**, de forma que no sea posible llegar a deducir la identidad de la persona a la cual corresponden los datos objeto de tratamiento, tras haberse asegurado el aislamiento y la no vinculación.

Existen muchas y muy variadas metodologías de anonimización, y es muy posible que sea necesario combinar varias de ellas para asegurar la completa disociación en un proyecto *Big Data*. En cualquier caso, debe tenerse en cuenta que la anonimización debe asegurarse de principio a fin, es decir, en todos y cada uno de los procesos y tratamientos de información específicos asociados al proyecto en cuestión, por lo que la selección de la mejor metodología de anonimización dependerá de las circunstancias particulares de cada proyecto *Big Data*.

Propiedad intelectual de las bases de datos

Existe otro aspecto importante relacionado con estos proyectos, y es el relativo a la propiedad, titularidad y protección jurídica del proyecto *Big Data* en su conjunto, como de los diferentes elementos intangibles que lo integran.

La normativa sobre propiedad intelectual incluye dos niveles de derechos de propiedad intelectual para las bases de datos:

- **La protección idéntica al resto de derechos de autor**, y que se confiere a las bases de datos cuya especial selección y disposición de los elementos que las integran puedan ser consideradas en sí mismas como creaciones intelectuales; y

- El derecho "*sui generis*", que protege la inversión sustancial que realiza el fabricante de la base de datos, ya sea en medios financieros, empleo de tiempo, esfuerzo, energía u otros similares, para la obtención, verificación o presentación de su contenido, con independencia de que la selección o disposición de sus elementos pueda ser considerada o no como una creación intelectual.

En un proyecto *Big Data*, ambos tipos de titularidad y propiedad intelectual pueden llegar a coexistir, y pueden darse situaciones de conflicto de titularidad. Por un lado, el derecho *sui generis* sobre las bases de datos resultantes de un proyecto *Big Data* corresponderá, casi siempre, al impulsor del proyecto, por ser quien realiza el esfuerzo de financiación de todos los recursos asociados al mismo. Pero ello puede no ser así respecto al esfuerzo intelectual creativo relativo a la especial selección y disposición de los datos asociados al proyecto *Big Data*, esfuerzo que, por aplicación del principio de autoría consagrado por la normativa de propiedad intelectual, puede corresponder a la empresa consultora externa que realice dicha aportación creativa al proyecto.

Por dicho motivo, si el impulsor de un proyecto *Big Data* quiere asegurarse de que le corresponden en exclusiva todos los derechos de propiedad intelectual sobre todos y cada uno de los elementos y resultados del mismo, será recomendable que asegure dicha titularidad por vía contractual, en particular, en relación con empresas consultoras externas que participen en el diseño y ejecución del proyecto.

Protección jurídica de los algoritmos

Resulta paradójico que los algoritmos, que constituyen uno de los elementos que puede aportar mayor creatividad a un proyecto *Big Data*, no dispongan de una protección jurídica específica.

Los algoritmos no pueden considerarse programas de ordenador, puesto que no están expresados en secuencias de instrucciones en lenguaje de programación. Los algoritmos constituyen la secuencia ordenada de operaciones que se pretende realizar para conseguir un resultado concreto o resolver un problema específico, y, por lo tanto, constituyen las ideas o principios en los que se basan los programas de ordenador asociados a los proyectos *Big Data*. Dichas ideas y principios están expresamente excluidos de ser considerados como propiedad intelectual.

Por otra parte, y conforme a diversas decisiones adoptadas por la Oficina Europea de Patentes y Marcas, los algoritmos tampoco reúnen las condiciones para ser protegidos como patente.

Por lo tanto, la única forma de protección posible para los algoritmos es la disponible para los **secretos industriales y comerciales**, sometiendo su conocimiento a obligación de confidencialidad (aplicable tanto al personal propio del impulsor del proyecto *Big Data*, como a los consultores y proveedores externos que participen en el mismo), y recurriendo a las acciones civiles y penales disponibles en nuestro ordenamiento en caso de acceso, apropiación o difusión no autorizados. En este ámbito específico de protección es de esperar que la reciente Directiva 2016/943 de Protección de los Secretos Comerciales, y su futura incorporación a nuestro derecho, venga a reforzar los recursos judiciales disponibles para la adecuada protección jurídica de los algoritmos que formen parte de los elementos integrantes de proyectos *Big Data*.

