Big Data y protección de datos [1]

Vicente Guasch Portas, Escola Universitària de Turisme del Consell d'Eivissa

Resumen

Big Data supone una gran revolución tecnológica en la actualidad. Cuando hablamos de Big Data nos referimos a la gestión y análisis de enormes volúmenes de datos que no pueden ser tratados de manera convencional, ya que superan los límites y capacidades de las herramientas de software habitualmente utilizadas para la captura, gestión y procesamiento de datos. El principal objetivo del *Big Data*, es la obtención de información que facilite la toma de decisiones, incluso en tiempo real.

Sin embargo, los beneficios generados por el *Big Data* también van acompañados de riesgos. En particular, podemos señalar el riesgo de este análisis masivo de datos sobre la protección de datos de carácter personal. En el *Big Data* podemos encontrar tratamientos que no sean lícitos por realizarse sin respaldo legal para ello, pero también existen riesgos y amenazas de la mano de aspectos técnicos y de seguridad.

Palabras clave: big data, datos personales, RGPD, protección de datos, anonimización.

Abstract:

Big Data is a great technological revolution today. When we talk about *Big Data* we refer to the management and analysis of huge volumes of data that can not be treated in a conventional manner, since they exceed the limits and capabilities of the software tools commonly used for data capture, management and processing. The main objective of *Big Data* is to obtain information that facilitates decision making, even in real time.

However, the benefits generated by *Big Data* are also accompanied by risks. In particular, we can point out the risk of this massive analysis of data on the protection of personal data. In *Big Data* we can find treatments that are not lawful because they are carried out without legal backing, but there are also risks and threats along with technical and security aspects.

Keywords: big data, personal data, GDPR, data protection, anonymisation.

Introducción

Es muy común hablar de las enormes ventajas que tiene el *Big Data* en la gestión empresarial. ¿Pero que es realmente el *Big Data*?

El *Big Data* se ha definido por Gartner[2] como *activos de información de gran volumen, velocidad y variedad, que exigen formas innovadoras y rentables de procesamiento de la información para mejorar la comprensión y la toma de decisiones*.

De acuerdo a esa definición, al *Big Data* se le relaciona con las tres *V* de volumen, velocidad y variedad. Por otra parte, el *Big Data* hace referencia a conjuntos de datos tan grandes que aplicaciones informáticas tradicionales no son suficientes para tratar con ellos. Son necesarios nuevos sistemas y procedimientos para la manipulación de esos bloques inmensos de información.

Como pone de manifiesto el Supervisor Europeo de Protección de Datos[3], el tratamiento de enormes cantidades de datos, si se hace de manera responsable, puede ofrecer importantes beneficios y eficiencias para la sociedad y los individuos en el campo de la salud, la investigación científica, el medio ambiente y otras áreas específicas. Pero hay una gran preocupación con el impacto real y potencial del procesamiento de cantidades masivas de datos sobre los derechos y las libertades de las personas, incluido su derecho a la privacidad.

Como señala la Agencia de protección de datos británica, algunos de los aspectos distintivos del *Big Data* son[4]:

- El uso de algoritmos.
- La opacidad de tratamiento.
- La tendencia a recoger *todos los datos*.
- La reutilización de los datos.
- El uso de nuevos tipos de datos.

Cuando los datos que son objeto de tratamiento en el *Big Data*, son datos personales, se deberá someter ese tratamiento a la normativa aplicable en materia de protección de datos.

La protección de las personas físicas en relación con el tratamiento de datos personales es un derecho fundamental. El artículo 8, apartado 1, de la Carta de los Derechos Fundamentales de la Unión Europea y el artículo 16, apartado 1, del Tratado de Funcionamiento de la Unión Europea establecen que toda persona tiene derecho a la protección de los datos de carácter personal que le conciernan.

En otro orden, el 25 de mayo de 2016 entró en vigor el Reglamento General de Protección de Datos[5] (en adelante, RGPD). Esta normativa reguladora de la protección de datos en la UE comenzó a aplicarse el 25 de mayo de 2018.

En el artículo 4.1) del RGPD se define como *datos personales* a *toda información sobre una persona física identificada o identificable". Y añade que "se considerará persona física identificable toda persona cuya identidad pueda determinarse, directa o indirectamente, en particular mediante un identificador, como por ejemplo un nombre, un número de identificación, datos de localización, un identificador en línea o uno o varios elementos propios de la identidad física, fisiológica, genética, psíquica, económica, cultural o social de dicha persona*.

El Grupo de Trabajo del Artículo 29 (GT29) analizó en su Dictamen 4/2007[6] los cuatro componentes principales que pueden distinguirse en la definición de *datos personales*, esto es: *toda información*, *sobre*, *persona física* e *identificada o identificable*. Estos cuatro componentes están estrechamente ligados y se complementan entre sí, pero juntos determinan si una determinada información debe ser, o no, considerada como *datos personales*.

La expresión *toda información* indica claramente la voluntad del legislador de dar un sentido amplio al concepto datos personales. Desde el punto de vista de la naturaleza de la información, el concepto de datos personales incluye todo tipo de afirmaciones sobre una persona. Por consiguiente, abarca información objetiva, pero también informaciones, opiniones o evaluaciones subjetivas. Desde el punto de vista del formato o el soporte en que la información está contenida, el concepto de datos personales incluye la información disponible en cualquier forma (alfabética, numérica, gráfica, fotográfica, sonora, etc.) y en cualquier soporte (en papel, en un disco duro, etc.).

El componente *sobre* de la definición es crucial. De modo general, se puede considerar que la información versa *sobre* una persona cuando se refiere a ella. Como señaló el GT29 en su documento de trabajo WP 105[7], un *dato se refiere a*

una persona si hace referencia a su identidad, sus características o su comportamiento o si esa información se utiliza para determinar o influir en la manera en que se la trata o se la evalúa.

El tercer componente exige que la protección se aplique a la *persona física*, es decir a un ser humano. El derecho a la protección de los datos personales es, en ese sentido, universal sin circunscribirse a los nacionales o residentes en determinado país.

El último componente exige que la información se refiera a una persona física *identificada o identificable*. De modo general, se puede considerar *identificada* a una persona física cuando, dentro de un grupo de personas, se la *distingue* de todos los demás miembros del grupo. Por contra, la persona física es *identificable* cuando, aunque no se la haya identificado todavía, sea posible hacerlo. Para determinar si una persona es identificable, hay que considerar el conjunto de los medios que puedan ser razonablemente utilizados por el responsable del tratamiento o por cualquier otra persona, para identificar a dicha persona.

En muchas ocasiones los datos no encajan dentro de la definición de *datos personales*. Así sucede, por ejemplo, cuando no puede afirmarse que los datos se refieren a una persona física, o cuando no cabe hablar de persona identificada o identificable. En este sentido, se puede pensar en el tratamiento de grandes volúmenes de datos generados por sensores para monitorear los fenómenos naturales o atmosféricos como el clima o la contaminación, o para el seguimiento de los aspectos técnicos de los procesos de fabricación. En estos y en otros muchos ámbitos se pueden emplear las técnicas de *Big Data* sin necesidad de utilizar datos personales.

Si la información que se está tratando no encaja en el concepto de *datos personales*, la consecuencia es que la normativa sobre protección de datos no es aplicable. Así lo confirma el considerando 26 del RGPD: *los principios de protección de datos no deben aplicarse a la información anónima, es decir información que no guarda relación con una persona física identificada o identificable, ni a los datos convertidos en anónimos de forma que el interesado no sea identificable, o deje de serlo. En consecuencia, el presente Reglamento no afecta al tratamiento de dicha información anónima ….*

Big Data y datos personales

Con mucha frecuencia, a través de *Big Data* se tratarán datos personales. Una de las principales utilidades del tratamiento de grandes volúmenes de datos por parte de las empresas y de los gobiernos reside en el potencial predictivo que se deriva de la supervisión del comportamiento humano, tanto individual como colectivo. En otros casos, las empresas utilizan grandes volúmenes de datos con el fin de ofrecer productos o servicios de una manera más eficiente o para proporcionar un servicio más personalizado.

Por lo tanto, si se tratan datos personales, habrá que tener en consideración lo establecido en la legislación sobre protección de datos.

En el artículo 5.1 del RGPD se exige que los datos personales sean:

- Tratados de manera lícita, leal y transparente.
- Recogidos con fines determinados, explícitos y legítimos, y no tratados ulteriormente de manera incompatible con dichos fines.
- Adecuados, pertinentes y limitados a lo necesario en relación con los fines para los que son tratados.
- Exactos y, si fuera necesario, actualizados.
- Mantenidos de forma que se permita la identificación de los interesados durante no más tiempo del necesario para los fines del tratamiento de los datos personales.

El tratamiento lícito

Para que el tratamiento sea lícito, los datos personales deben ser tratados con el consentimiento del interesado o sobre alguna otra base legítima establecida conforme a Derecho[8]. De acuerdo al artículo 6.1 del RGPD, el tratamiento solo será lícito si se cumple al menos una de las siguientes condiciones:

- El interesado dio su consentimiento para el tratamiento de sus datos personales para uno o varios fines específicos.
- El tratamiento es necesario para la ejecución de un contrato en el que el interesado es parte o para la aplicación a petición de este de

medidas precontractuales.

- El tratamiento es necesario para el cumplimiento de una obligación legal aplicable al responsable del tratamiento.
- El tratamiento es necesario para proteger intereses vitales del interesado o de otra persona física.
- El tratamiento es necesario para el cumplimiento de una misión realizada en interés público o en el ejercicio de poderes públicos conferidos al responsable del tratamiento.
- El tratamiento es necesario para la satisfacción de intereses legítimos perseguidos por el responsable del tratamiento o por un tercero, siempre que sobre dichos intereses no prevalezcan los intereses o los derechos y libertades fundamentales del interesado que requieran la protección de datos personales.

Puede hacerse una distinción entre el caso de que los datos personales se traten sobre la base del consentimiento[9] y los cinco casos restantes. Estos últimos describen supuestos en los que el tratamiento puede ser necesario en un contexto específico. Pero en el caso del consentimiento son los mismos interesados los que autorizan el tratamiento de sus datos personales. La decisión de permitir que sus datos sean tratados depende de ellos. En otras palabras, en el caso del consentimiento el motivo de legitimación se centra en la libre determinación del interesado.

El consentimiento debe darse mediante un acto afirmativo claro que refleje una manifestación de voluntad libre, específica, informada, e inequívoca del interesado de aceptar el tratamiento de datos de carácter personal que le conciernen. Además, el consentimiento debe darse para todas las actividades de tratamiento realizadas con el mismo o los mismos fines. Cuando el tratamiento tenga varios fines, debe darse el consentimiento para todos ellos[10]. Pero en el contexto del *Big Data* esta es una exigencia que no es práctica. El análisis de grandes volúmenes de datos tiene muchas veces un carácter experimental, en el que la búsqueda de nuevos usos para los datos es muy habitual.

Para que el tratamiento sea lícito, si no se puede usar el consentimiento de las personas afectadas, el *Big Data* deberá usar como alternativa en la mayor parte de las ocasiones, el último de los puntos relacionados como herramienta de legitimación: el tratamiento es necesario para la satisfacción de intereses legítimos perseguidos por el responsable del tratamiento o por un tercero, siempre que sobre dichos intereses no prevalezcan los intereses o los derechos y libertades fundamentales del interesado que requieran la protección de datos personales. En este caso el interés legítimo del responsable del tratamiento o de un tercero debe sopesarse en relación con los intereses o los derechos y libertades fundamentales del interesado. El resultado de esta prueba determinará en gran medida si puede, o no, considerarse un fundamento jurídico del tratamiento.

Esta prueba no consiste simplemente en ponderar dos *pesos* fácilmente cuantificables y comparables. Por el contrario, puede exigir una compleja evaluación que tenga en consideración todos los factores en juego. El GT29 ha estudiado en profundidad como llevar a cabo dicha evaluación en su dictamen 06/2014[11]. Para el GT29, el interés legítimo no deberá considerarse un fundamento jurídico que solo puede utilizarse con moderación para cubrir las lagunas en situaciones raras o imprevistas como un último recurso si no se pueden aplicar otros motivos de legitimación. Tampoco deberá percibirse como una opción preferente ni deberá extenderse su uso indebidamente porque se considere menos restrictiva que los demás fundamentos jurídicos. Por el contrario, se trata de un medio tan válido como cualquier otro para legitimar el tratamiento de datos personales. El GT29 aspira a un enfoque equilibrado que garantice la flexibilidad necesaria a los responsables del tratamiento de datos en situaciones en las que no exista un impacto indebido sobre los interesados.

Fines determinados, explícitos y legítimos

Habíamos indicado que los datos deben ser recogidos con fines determinados, explícitos y legítimos, y no tratados ulteriormente de manera incompatible con dichos fines.

El concepto de limitación de la finalidad juega un papel crucial en la aplicación de la normativa sobre

protección de datos, ya que constituye un requisito previo para otros requisitos de calidad de datos. Contribuye a la transparencia, la seguridad jurídica y la previsibilidad y su objetivo es proteger a los titulares de los datos mediante el establecimiento de límites sobre cómo serán utilizados sus datos.

La limitación de la finalidad puede constituir una barrera para el desarrollo del *Big Data*. El análisis de grandes volúmenes de datos usando múltiples algoritmos permite revelar correlaciones inesperadas que pueden llevar a que los datos sean usados para nuevos propósitos. La limitación de la finalidad limita la libertad para llevar a cabo estos estudios, fuente en muchas ocasiones de nuevos descubrimientos e innovaciones.

El RGPD exige en su artículo 30 que cada responsable lleve un registro de las actividades de tratamiento efectuadas bajo su responsabilidad. Dicho registro deberá contener, entre otros datos, los fines del tratamiento. Dada la mecánica de funcionamiento del *Big Data*, esta obligación puede ser de difícil cumplimiento.

El concepto de finalidad tiene dos bloques principales: los datos personales deben ser recogidos con fines determinados, explícitos y legítimos (especificación de los fines) y no pueden ser tratados posteriormente de manera incompatible con dichos fines (uso compatible).

Sobre la especificación de los fines, en el GT29 se destaca las siguientes consideraciones fundamentales[12]:

- Los fines han de ser específicos. Esto significa que antes, o en el momento en que se produce la recogida de datos de carácter personal, los fines deben ser identificados con precisión.
- Los fines han de ser explícitos, es decir, claramente revelados, con el fin de asegurarse de que todo el mundo tiene la misma comprensión inequívoca de los fines del tratamiento.
- Los fines han de ser legítimos. La legitimidad es un requisito amplio, que va más allá de lo exigido por la normativa de protección de datos. También se extiende a otras áreas del derecho.

En cuanto al uso compatible de los datos, el GT29 entiende que un tratamiento adicional no debe ser incompatible con los fines para los que se recogieron los datos personales. La prohibición de uso incompatible establece entonces una limitación en su uso posterior. Se requiere que se haga una distinción entre el uso adicional que sea compatible, y el uso posterior que sea incompatible, y por lo tanto, prohibido.

Al prohibir la incompatibilidad en lugar de requerir la compatibilidad, el legislador parece dar cierta flexibilidad con respecto a su uso posterior. El procesamiento adicional para un propósito diferente significa no necesariamente y de forma automática que es incompatible. Deberá evaluarse caso por caso la compatibilidad.

En el mismo sentido se expresa el considerando 50 del RGPD: *Con objeto de determinar si el fin del tratamiento ulterior es compatible con el fin de la recogida inicial de los datos personales, el responsable del tratamiento, tras haber cumplido todos los requisitos para la licitud del tratamiento original, debe tener en cuenta, entre otras cosas, cualquier relación entre estos fines y los fines del tratamiento ulterior previsto, el contexto en el que se recogieron los datos, en particular las expectativas razonables del interesado basadas en su relación con el responsable en cuanto a su uso posterior, la naturaleza de los datos personales, las consecuencias para los interesados del tratamiento ulterior previsto y la existencia de garantías adecuadas tanto en la operación de tratamiento original como en la operación de tratamiento ulterior prevista*.

Datos adecuados, pertinentes y limitados

Los datos personales serán adecuados, pertinentes y limitados a lo necesario en relación con los fines para los que son tratados (minimización de datos). Es decir, las organizaciones deben reducir al mínimo la cantidad de datos que recogen y tratan, así como el tiempo que mantienen esos datos.

Big Data puede dar lugar a una recogida de datos personales que sea excesiva para el propósito del tratamiento. Además, las organizaciones pueden conservar los datos personales más de lo necesario, ya que las aplicaciones informáticas de *Big Data* son capaces de analizar grandes volúmenes de datos.

Big Data tiende a implicar la recogida y análisis de tantos datos como sea posible. Al referirse a la minimización de datos no se hace referencia solo a

la cantidad de datos que se utiliza, sino a si son necesarios para los fines del tratamiento, o si son excesivos. La recogida excesiva de datos que no se llegan a utilizar es un inconveniente real en buena parte de las empresas europeas. Esa recopilación de datos excesiva supone un problema de protección de datos, pero también hace que sea más difícil localizar y trabajar sobre los datos que realmente necesitan las empresas[13].

De acuerdo al GT29 en su documento de trabajo WP 221[14], los principios de limitación de la finalidad y la minimización de los datos se presentan como principales preocupaciones en el tratamiento de grandes volúmenes de datos de carácter personal.

Como señala Elena Gil[15], en el caso del *Big Data*, el principio de minimización de datos no se cumple en la práctica. Este principio implica que los datos recopilados no deben ser excesivos, pero esto se contrapone contra la misma lógica del *Big Data*. Los nuevos modelos analíticos se basan precisamente en el estudio de cantidades masivas de datos sin los cuales no podría extraerse el conocimiento que nos permite el *Big Data*.

Datos exactos y, si fuera necesario, actualizados

Los datos personales deben ser exactos y, cuando sea necesario, actualizados. Esto es obviamente una buena práctica en términos de gestión de la información, pero también está íntimamente relacionado con los derechos de la persona. Las personas tienen derecho a que se rectifiquen los datos inexactos.

Por el contrario, se ha sugerido que, en determinadas circunstancias, el tratamiento de datos masivos puede tolerar una cierta cantidad de datos inexactos, ya que los volúmenes de datos que están siendo procesados son generalmente enormes. Según indica la I.C.O., *un cierto nivel de* desorden*, como el nombre o la dirección incorrecta, puede no ser un problema cuando los análisis se utilizan para detectar tendencias generales. Pero es mucho más probable que sea problemático cuando el procesamiento se utiliza para perfilar los individuos particulares*[16].

Todo interesado tendrá derecho a no ser objeto de una decisión basada únicamente en el tratamiento automatizado, incluida la elaboración de perfiles, que produzca efectos jurídicos en él o le afecte significativamente de modo similar[17].

En el perfilado de la gente, cuando los datos son inexactos pueden dar lugar a predicciones incorrectas sobre su comportamiento o su salud, la solvencia o el riesgo[18].

El considerando 39 del RGPD se expresa con toda claridad: *Deben tomarse todas las medidas razonables para garantizar que se rectifiquen o supriman los datos personales que sean inexactos*. El artículo 5 del RGPD añade además que esa rectificación o supresión deben efectuarse *sin dilación*.

Y el artículo 16 del RGPD regula el derecho de rectificación: *El interesado tendrá derecho a obtener sin dilación indebida del responsable del tratamiento la rectificación de los datos personales inexactos que le conciernan. Teniendo en cuenta los fines del tratamiento, el interesado tendrá derecho a que se completen los datos personales que sean incompletos, inclusive mediante una declaración adicional*.

Conservación de datos por el tiempo necesario

El artículo 5 del RGPD exige que los datos sean mantenidos de forma que se permita la identificación de los interesados durante no más tiempo del necesario para los fines del tratamiento de los datos personales.

Para garantizar que los datos personales no se conservan más tiempo del necesario, el responsable del tratamiento ha de establecer plazos para su supresión o revisión periódica[19].

Cuando no se mantienen los datos personales más de lo necesario para la finalidad para la que se está procesando, se está protegiendo la privacidad de los datos de las personas. Sin embargo, en el mundo del *Big Data* esto puede ser cuestionado por dos razones. En primer lugar, la capacidad de almacenar datos aumenta continuamente y el coste de almacenamiento disminuye con rapidez. En segundo lugar, la capacidad de análisis del *Big Data* para procesar grandes volúmenes de datos puede animar a los responsables del tratamiento a mantener largas series de datos históricos más allá del tiempo necesario para fines comerciales

normales.

Sin embargo, las organizaciones deben ser capaces de articular desde el principio por qué tienen que recoger y procesar conjuntos de datos particulares. Tienen que tener claro lo que esperan obtener mediante el procesamiento de los datos, y por lo tanto asegurarse de que son pertinentes y no excesivos en relación con ese objetivo. El reto es definir los fines del tratamiento y determinar qué datos serán relevantes.

Muy relacionado con el plazo de conservación de los datos, en el artículo 17 del RGPD se regula *el derecho al olvido* (derecho de supresión). El interesado tendrá derecho a obtener del responsable del tratamiento la supresión de los datos personales que le conciernan en determinadas circunstancias. Entre ellas podemos destacar:

- Cuando los datos personales ya no sean necesarios en relación con los fines para los que fueron recogidos.
- Cuando el interesado retire el consentimiento.
- Cuando los datos personales hayan sido tratados ilícitamente.

Cuando haya hecho públicos los datos personales y esté obligado a suprimir dichos datos, el responsable del tratamiento, teniendo en cuenta la tecnología disponible y el coste de su aplicación, adoptará medidas razonables, incluidas medidas técnicas, con miras a informar a los responsables que estén tratando los datos personales, de la solicitud del interesado de supresión de cualquier enlace a esos datos personales, o cualquier copia o réplica de los mismos[20].

Por otra parte, de acuerdo al artículo 15 del RGPD, en donde se regula el derecho de acceso del interesado, el interesado tendrá derecho a obtener del responsable del tratamiento, entre otras, la siguiente información: *de ser posible, el plazo previsto de conservación de los datos personales o, de no ser posible, los criterios utilizados para determinar este plazo*.

Inaplicación de la normativa sobre protección de datos en el *Big Data*

El RGPD se aplica al tratamiento total o parcialmente automatizado de datos personales, así como al tratamiento no automatizado de datos personales contenidos o destinados a ser incluidos en un fichero. Tal como ya habíamos indicado previamente, se entiende por *datos personales* toda información sobre una persona física identificada o identificable.

Para determinar si una persona física es identificable, deben tenerse en cuenta todos los medios, que razonablemente pueda utilizar el responsable del tratamiento o cualquier otra persona para identificar directa o indirectamente a la persona física. Para determinar si existe una probabilidad razonable de que se utilicen medios para identificar a una persona física, deben tenerse en cuenta todos los factores objetivos, como los costes y el tiempo necesarios para la identificación, teniendo en cuenta tanto la tecnología disponible en el momento del tratamiento como los avances tecnológicos.

Incluso los datos personales seudonimizados[21], que cabría atribuir a una persona física mediante la utilización de información adicional, deben considerarse información sobre una persona física identificable. Pero la aplicación de la seudonimización a los datos personales puede reducir los riesgos para los interesados afectados y ayudar a los responsables y a los encargados del tratamiento a cumplir sus obligaciones de protección de los datos[22].

En la misma línea, el artículo 32.1 del RGPD exige al responsable y al encargado del tratamiento la aplicación de medidas técnicas y organizativas apropiadas para garantizar un nivel de seguridad adecuado al riesgo, que en su caso incluya, entre otras posibilidades, la seudonimización y el cifrado de datos personales.

Pero los principios de protección de datos no deben aplicarse a la información anónima, es decir información que no guarda relación con una persona física identificada o identificable, ni a los datos convertidos en anónimos de forma que el interesado no sea identificable, o deje de serlo[23].

Conforme aumentan los volúmenes de información y los tipos de datos que generan los dispositivos electrónicos, los sensores y las redes y se reduce el coste de almacenamiento hasta cantidades insignificantes, crecen el interés de los ciudadanos y la demanda de reutilización de estos

datos. Unos *datos abiertos* pueden aportar beneficios visibles a la sociedad, a las personas y a las organizaciones, pero solo si se respetan los derechos de todos a la protección de los datos personales y a la vida privada.

La anonimización[24] puede ser una buena estrategia en el *Big Data* para obtener estos beneficios al mismo tiempo que se mitigan los riesgos. Cuando un conjunto de datos se anonimiza realmente y no es posible ya identificar a las personas, no es aplicable la legislación sobre protección de datos[25].

No obstante, la generación de un conjunto de datos verdaderamente anónimo a partir de un gran conjunto de datos personales, conservando al mismo tiempo la información subyacente que se requiere para llevar a cabo la tarea, no es un propósito sencillo. Puede que no sea posible establecer con absoluta certeza que un individuo no podrá ser identificado a partir de los datos anonimizados o en conjunto con otros datos distintos. Los responsables de tratamiento deben esforzarse para minimizar los riesgos de reidentificación.

El GT29 ha adoptado un dictamen sobre técnicas de anonimización[26] en el que analiza la eficacia y las limitaciones de las técnicas de anonimización existentes, atendiendo al marco legal de la UE sobre protección de datos, y formula recomendaciones para la gestión de estas técnicas teniendo en cuenta el riesgo residual de identificación inherente a cada una de ellas. Se reconoce el valor potencial de la anonimización, pero muestra la dificultad de crear un conjunto de datos verdaderamente anónimo conservando, sin embargo, toda la información subyacente requerida para la tarea.

Para el GT29, el resultado de la anonimización, entendida esta como una técnica aplicada a los datos personales, debe ser, de acuerdo con el actual estado de la tecnología, tan permanente como el borrado. En otras palabras: debe garantizarse que es imposible tratar los datos personales.

El análisis de las referencias a la anonimización en la normativa sobre protección de datos permite poner de manifiesto cuatro características fundamentales:

- La anonimización puede ser el resultado de un tratamiento de datos personales realizado para impedir de forma irreversible la identificación del interesado.
- Pueden considerarse varias técnicas de anonimización, sin que la legislación europea contenga ninguna norma prescriptiva.
- Hay que dar importancia a los elementos contextuales: debe considerarse *el conjunto de los medios que puedan ser razonablemente utilizados* para la identificación por parte del responsable del tratamiento o de un tercero, prestando especial atención a lo que se entiende, en el estado actual de la técnica, como *medios que puedan ser razonablemente utilizados* (dado el incremento de la potencia de los ordenadores y de las herramientas disponibles).
- La anonimización lleva implícito un factor de riesgo que ha de tenerse en cuenta al evaluar la validez de las técnicas de anonimización, incluidos los posibles usos de los datos *anonimizados* mediante estas, además de considerarse asimismo la gravedad y probabilidad del riesgo.

Para la AEPD[27], los procesos de anonimización son una herramienta válida para garantizar la privacidad de los datos personales y sus limitaciones son inherentes al avance de la tecnología. Existe una proporcionalidad manifiesta en lo que respecta a la capacidad tecnológica de anonimizar y la posibilidad de la reidentificación de las personas cuyos datos han sido anonimizados, es decir, la misma capacidad de la tecnología para anonimizar datos personales puede ser utilizada para la reidentificación de las personas[28].

No es posible considerar que los procesos de anonimización garanticen con total certeza la no reidentificación de las personas, por lo que será necesario sustentar la fortaleza de la anonimización en otras medidas que sirvan tanto para atenuar los riesgos de reidentificación de las personas como para paliar las consecuencias de que éstos se materialicen.

Big Data y otras herramientas para la protección de datos

De acuerdo al artículo 35 del RGPD, cuando sea

probable que un tipo de tratamiento, en particular si utiliza nuevas tecnologías, por su naturaleza, alcance, contexto o fines, entrañe un alto riesgo para los derechos y libertades de las personas físicas, el responsable del tratamiento deberá realizar, antes del tratamiento, una *evaluación del impacto*, de las operaciones de tratamiento en la protección de datos personales[29].

Por las características del *Big Data*, es muy probable que la evaluación del impacto en la protección de datos se convierta en un requisito imprescindible.

La evaluación deberá incluir como mínimo:

a. una descripción sistemática de las operaciones de tratamiento previstas y de los fines del tratamiento, inclusive, cuando proceda, el interés legítimo perseguido por el responsable del tratamiento;
b. una evaluación de la necesidad y la proporcionalidad de las operaciones de tratamiento con respecto a su finalidad;
c. una evaluación de los riesgos para los derechos y libertades de los interesados, y
d. las medidas previstas para afrontar los riesgos, incluidas garantías, medidas de seguridad y mecanismos que garanticen la protección de datos personales, y a demostrar la conformidad con el RGPD, teniendo en cuenta los derechos e intereses legítimos de los interesados y de otras personas afectadas.

El resultado de la evaluación debe tenerse en cuenta cuando se decidan las medidas adecuadas que deban tomarse con el fin de demostrar que el tratamiento de los datos personales es conforme con el RGPD.

Si una evaluación de impacto relativa a la protección de datos muestra que las operaciones de tratamiento entrañan un alto riesgo que el responsable no puede mitigar con medidas adecuadas en términos de tecnología disponible y costes de aplicación, debe consultarse a la autoridad de control antes del tratamiento (considerando 84 y artículo 36 del RGPD).

Otra obligación distinta a la que tendrán que someterse la mayor parte de operaciones de *Big Data* viene recogida en el artículo 25 del RGPD. Se trata de la *protección de datos desde el diseño y por defecto*. Teniendo en cuenta el estado de la técnica, el coste de la aplicación y la naturaleza, ámbito, contexto y fines del tratamiento, así como los riesgos de diversa probabilidad y gravedad que entraña el tratamiento para los derechos y libertades de las personas físicas, el responsable del tratamiento deberá aplicar, tanto en el momento de determinar los medios de tratamiento como en el momento del propio tratamiento, medidas técnicas y organizativas apropiadas, como la seudonimización, concebidas para aplicar de forma efectiva los principios de protección de datos, como la minimización de datos, e integrar las garantías necesarias en el tratamiento, a fin de cumplir los requisitos del RGPD y proteger los derechos de los interesados.

El responsable del tratamiento aplicará las medidas técnicas y organizativas apropiadas con miras a garantizar que, por defecto, solo sean objeto de tratamiento los datos personales que sean necesarios para cada uno de los fines específicos del tratamiento. Esta obligación se aplicará a la cantidad de datos personales recogidos, a la extensión de su tratamiento, a su plazo de conservación y a su accesibilidad.

Conclusiones

El tratamiento de grandes volúmenes de datos, si se hace de manera responsable, puede ofrecer importantes beneficios para la sociedad y los individuos. Pero hay una gran preocupación con el impacto del *Big Data* sobre los derechos y las libertades de las personas, incluido su derecho a la protección de datos.

Los modelos de negocio que explotan las nuevas capacidades para la recogida masiva de datos, la transmisión instantánea, la combinación y la reutilización de la información personal para fines no previstos, exigen un examen a fondo sobre la forma en que se aplica la normativa sobre protección de datos.

No se puede prohibir el *Big Data*, ya que el coste económico sería enorme. Pero si los responsables de tratamiento efectúan grandes inversiones para encontrar maneras innovadoras de hacer uso de los datos personales, también deben realizar grandes esfuerzos en la correcta aplicación de la legislación sobre protección de datos.

Los principios de limitación de la finalidad y la minimización de los datos se presentan como principales dificultades de cumplimiento en el *Big Data*. Los responsables de tratamiento solo deberían recoger datos personales con fines determinados, explícitos y legítimos, y no los tendrían que tratar posteriormente de manera incompatible con dichos fines. Los datos personales también deberían ser adecuados, pertinentes y no excesivos en relación con los fines para los que se recojan.

La anonimización de los datos personales puede ser una herramienta muy útil en el *Big Data*. A través de ella se puede llevar a cabo el tratamiento de datos sin menoscabar el respeto a la protección de datos.

A través de la anonimización se quiere eliminar la posibilidad de identificación de las personas, si bien el avance de la tecnología no permite garantizar el anonimato absoluto, especialmente a lo largo del tiempo. Aun así, a través de la anonimización se pueden ofrecer mayores garantías de privacidad a las personas.

Una *evaluación del impacto*, de las operaciones de tratamiento en la protección de datos personales y la *protección de datos desde el diseño y por defecto* pueden ser otras herramientas imprescindibles para que el *Big Data* minimice el riesgo del tratamiento de los datos de carácter personal.

Notas

[1] El presente trabajo se ha elaborado en el marco del Proyecto Big Data*, Cloud Computing y otros nuevos retos jurídicos planteados por las tecnologías emergentes; en particular, su incidencia en el sector turístico* (DER2015-63595-R MINECO/FEDER), Investigadora Principal: Apol·lònia MARTÍNEZ NADAL, financiado por la Dirección General de Investigación, del Ministerio de Economía y Competitividad del Gobierno de España.

[2] Gartner IT glossary. Disponible en: **http://www.gartner.com/it-glossary**

[3] Dictamen 7/2015 del Supervisor Europeo de Protección de Datos, titulado *Meeting the challenges of big data*. Disponible en: **https://edps.europa.eu/sites/edp/files/publication/15-11-19_big_data_en.pdf**

[4] Así lo señala la I.C.O. británica en la página 9 de su documento *Big data, artificial intelligence, machine learning and data protection*. Disponible en: **https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf**

[5] REGLAMENTO (UE) 2016/679 DEL PARLAMENTO EUROPEO Y DEL CONSEJO, de 27 de abril de 2016, 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 por el que se deroga la Directiva 95/46/CE (Reglamento general de protección de datos).

[6] Dictamen 4/2007 (WP 136), *On the concept of personal data*, adoptado el 20 de junio de 2007 por el Grupo de Trabajo previsto en el artículo 29 de la Directiva 95/46/CE. Disponible en: **https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf**

[7] WP 105, *Working document on data protection issues related to RFID technology*, adoptado el 19 de enero de 2005 por el GT29. Disponible en: **https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2005/wp105_en.pdf**

[8] Así lo pone de manifiesto el considerando 40 del RGPD.

[9] Véase el dictamen 15/2011 (WP 187), *On the definition of consent*, adoptado el 13 de julio de 2011 por el GT29. Disponible en: **https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2011/wp187_en.pdf**

[10] Considerando 32 de RGPD.

[11] Dictamen 6/2014 (WP 217), *On the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC*, adoptado el 9 de abril de 2014 por el GT29. Disponible en: **https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp217_en.pdf**

[12] Dictamen 3/2013 (WP 203), *On purpose limitation*, adoptado el 2 de abril de 2013 por el GT29. Disponible en: **https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf**

[13] Así lo manifiesta la I.C.O. británica en la página 40 del documento ya citado *Big data, artificial intelligence, machine learning and data protection*.

[14] Declaración del GT29 *Statement on Statement of the WP29 on the impact of the development of big data on the protection of individuals with regard to the processing of their personal data in the EU* (WP 221), adoptada el 16 de septiembre de 2014. Disponible en: **https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp221_en.pdf**

[15] GIL GONZÁLEZ, E: *Big data, privacidad y protección de datos*, Imprenta del B.O.E., Madrid, 2016, p. 52.

[16] Así lo manifiesta la I.C.O. británica en la página 43 del documento ya citado *Big data, artificial intelligence, machine learning and data protection*.

[17] Artículo 22.1 del RGPD.

[18] De acuerdo al considerando 71 del RGPD *El interesado debe tener derecho a no ser objeto de una decisión, que puede incluir una medida, que evalúe aspectos personales relativos a él, y que se base únicamente en el tratamiento automatizado y produzca efectos jurídicos en él o le afecte significativamente de modo similar, como la denegación automática de una solicitud de crédito en línea o los servicios de contratación en red en los que no medie intervención humana alguna. Este tipo de tratamiento incluye la elaboración de perfiles consistente en cualquier forma de tratamiento de los datos personales que evalúe aspectos personales relativos a una persona física, en particular para analizar o predecir aspectos relacionados con el rendimiento en el trabajo, la situación económica, la salud, las preferencias o intereses personales, la fiabilidad o el comportamiento, la situación o los movimientos del interesado, en la medida en que produzca efectos jurídicos en él o le afecte significativamente de modo similar*.

[19] Considerando 39 del RGPD.

[20] Artículo 17.2 del RGPD.

[21] En el artículo 4 del RGPD se define la *seudonimización* como 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.

[22] Considerando 28 del RGPD.

[23] Considerando 26 del RGPD.

[24] Definición contenida en el diccionario de la R.A.E. sobre *Anonimizar*: Expresar un dato relativo a entidades o personas, eliminando la referencia a su identidad.

[25] En el ámbito del sector público, véase el documento de la AEPD titulado *Orientaciones sobre protección de datos en la REUTILIZACIÓN de la información del sector público*. Disponible en: **https://datos.gob.es/sites/default/files/doc/file/orientaciones_proteccion_datos_reutilizacion.pdf**

[26] Dictamen 5/2014 (WP 216), *On Anonymisation Techniques*, adoptado el 10 de abril de 2014 por el GT29. Disponible en: **https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf**

[27] Documento de la AEPD titulado *Orientaciones y garantías en los procedimientos de ANONIMIZACIÓN de datos personales*. Disponible en: **https://www.aepd.es/media/guias/guia-orientaciones-procedimientos-anonimizacion.pdf**

[28] Véase en este sentido el documento de la I.C.O. británica *Anonymisation: managing data protection risk - code of practice*. Disponible en: **https://ico.org.uk/media/1061/anonymisation-code.pdf**

[29] Véase el documento de la AEPD titulado *Guía práctica para las evaluaciones de impacto en la protección de los datos sujetas al RGPD*. Disponible en: **https://www.aepd.es/media/guias/guia-evaluaciones-de-impacto-rgpd.pdf**

Bibliografía

1. Monografías:

Cristea Uivaru, L. (2018).La protección de datos de carácter sensible: Historia Clínica Digital y *Big Data* en Salud. Bosch : Barcelona.

Garriga Domínguez, A. (2016). *Nuevos retos para la protección de datos personales. En la Era del* Big Data *y de la computación ubicua*. Dykinson : Madrid.

Gil González, E. (2016). *Big data, privacidad y protección de datos*. Boletín Oficial del Estado : Madrid.

Gudín Rodríguez Magariños, F. (2018). *Nuevo Reglamento Europeo de protección de Datos versus* Big Data. Tirant lo Blanch : Valencia.

Hoffmann-Riem, W. (2018). Big Data. *Desafíos también para el Derecho*. Civitas : Navarra.

Puyol Montero, J. (2015). *Aproximación Jurídica y Económica al* Big Data. Tirant lo Blanch : Valencia.

2. Documentos adoptados por el Grupo de Trabajo del artículo 29 de la Directiva:

WP 105 *Working document on data protection issues related to RFID technology*, adoptado el 19 de enero de 2005. Disponible en: **https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2005/wp105_en.pdf**

WP 136 *On the concept of personal data*, adoptado el 20 de junio de 2007. Disponible en: **https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf**

WP 187 *On the definition of consent*, adoptado el 13 de julio de 2011.

Disponible en: **https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2011/wp187_en.pdf**

WP 203 *On purpose limitation*, adoptado el 2 de abril de 2013. Disponible en:

https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf

WP 216 *On Anonymisation Techniques*, adoptado el 10 de abril de 2014. Disponible en: **https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf**

WP 217 *On the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC*, adoptado el 9 de abril de 2014. Disponible en:

https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp217_en.pdf

WP221 *Statement on Statement of the WP29 on the impact of the development of big data on the protection of individuals with regard to the processing of their personal data in the EU*, adoptado el 16 de septiembre de 2014. Disponible en: **https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp221_en.pdf**

3. Documentos de la Agencia Española de Protección de Datos:

Guía práctica para las evaluaciones de impacto en la protección de los datos sujetas al RGPD. Disponible en: **https://www.aepd.es/media/guias/guia-evaluaciones-de-impacto-rgpd.pdf**

Orientaciones sobre protección de datos en la REUTILIZACIÓN de la información del sector público. Disponible en: **https://datos.gob.es/sites/default/files/doc/file/orientaciones_proteccion_datos_reutilizacion.pdf**

Orientaciones y garantías en los procedimientos de ANONIMIZACIÓN de datos personales. Disponible en: **https://www.aepd.es/media/guias/guia-orientaciones-procedimientos-anonimizacion.pdf**

4. Documento del Supervisor Europeo de Protección de Datos:

Dictamen 7/2015 *Meeting the challenges of big data*. Disponible en: **https://edps.europa.eu/sites/edp/files/publication/15-11-19_big_data_en.pdf**

5. Documentos del Information Commissioner's Office (ICO):

Big data, artificial intelligence, machine learning and data protection. Disponible en: **https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf**

Anonymisation: managing data protection risk - code of practice. Disponible en: **https://ico.org.uk/media/1061/anonymisation-code.pdf**

Fecha de consulta de todos los enlaces electrónicos de este trabajo 30 de marzo de 2019

ISSN 2695-5334