CÓDIGO DE BUENAS PRÁCTICAS EN PROTECCIÓN DE DATOS PARA PROYECTOS

BIG DATA

AGENCIA ESPAÑOLA DE PROTECCIÓN DE DATOS

isms forum spain

CÓDIGO DE BUENAS PRÁCTICAS EN PROTECCIÓN DE DATOS PARA PROYECTOS

BIG DATA

AGENCIA ESPAÑOLA DE PROTECCIÓN DE DATOS

isms forum spain

Con la colaboración especial de:

dpi Data Privacy Institute

CSAES cloud security SPAIN alliance℠

autopistas an Abertis company

CaixaBank

ecix

FCC Servicios Ciudadanos

Hewlett Packard Enterprise

HUAWEI

MAPFRE

orange™

Telefónica

Universidad Politécnica Madrid

Coordinado por:
Carlos Alberto Sáiz

Aced, Emilio
del Álamo, José María
Armentia, Paula
Brito, Noemí
Buezo, Luis
Colom, José Luis
Cordón, Concepción
Corredera, Rosa
Díaz, Manuel
Díaz, Pablo
Fernández, Cecilia
García, Daniel
González, Francisco
Grifoll, Luis Estaban
Heras, María Rosario
Laredo, Jorge
Martín, Yod Samuel
Monleón, José Ramón
Mora, Elena
Muñoz, Antonio
Ortiz, Paula
Pantoja, Miguel Ángel
Pelegrín, Iolanda
Pérez, David
Sánchez, Óscar
Saracíbar, Esmeralda
Torrero, Juan Antonio

ÍNDICE

I.¿QUÉ ES EL BIG DATA?

I.1.- SIGNIFICADO, USOS Y TECNOLOGÍAS DEL BIG DATA.

Existen múltiples definiciones de Big Data de diversas fuentes. En síntesis, con dicho término se hace referencia al conjunto de tecnologías, algoritmos y sistemas empleados para recolectar datos a una escala y variedad no alcanzada hasta ahora y a la extracción de información de valor mediante sistemas analíticos avanzados soportados por computación en paralelo.

Al Big Data frecuentemente se le caracteriza mediante tres 'v': Volumen, Variedad y Velocidad:

• **Volumen** es la característica más obvia y que recoge el propio nombre de Big Data. Se pasa de manejar magnitudes de megabytes, gigabytes, como mucho Terabytes, a manejar Petabytes de forma cada vez más frecuente.

• Además de volumen de datos, su **Variedad** ha crecido exponencialmente, tanto por la tipología de datos como por sus fuentes. Se ha pasado de manejar datos estructurados en bases de datos procedentes, en su mayoría, de fuentes internas, a tratar datos estructurados, semiestructurados y desestructurados; de ser datos cuasi estáticos a datos dinámicos o en continuo cambio; de originarse en un número de fuentes limitadas a proceder de personas, máquinas, sensores, etc. Esta variedad y volumen, requieren un tratamiento diferente para poder convertirse en información.

• El tiempo es clave así que la **Velocidad** es la tercera 'v'. La captura, movimiento y proceso de los datos se hace a gran velocidad, llegando a ser en tiempo real en algunos casos.

Además, algunos autores y organizaciones han añadido nuevas 'v' para definir de forma más precisa al Big Data, por ejemplo, Veracidad (la calidad de los datos capturados es clave), Variabilidad (el significado de los datos cambia frecuentemente y se pueden producir inconsistencias que se han de manejar) y Valor (los ingresos o beneficios del Big Data).

Otro concepto relacionado que se maneja es el de *data lake* o lago de datos[1], en tanto que no sólo se trata de un almacenamiento de propósito específico de bajo coste y gran volumen, sino que se eleva a una agrupación o conglomerado de datos compartida por toda la organización en la que todo tipo de datos son accesibles simultáneamente por una variedad de motores de análisis sin apenas fricción. Es fácil ver que un concepto con un potencial tan grande para tratar volúmenes ingentes de datos –muchos de ellos personales- y desarrollar inferencias y correlaciones, lleva aparejados enormes posibilidades de progreso y a la vez retos importantes para la privacidad y la protección de datos personales a los que hay que hacer frente.

1 Un lago de datos o *data lake* es un repositorio de almacenamiento a gran escala que además proporciona una gran potencia de cómputo o procesamiento. En él se almacenan cualquier tipo de datos y tiene la potencialidad de gestionar una cantidad prácticamente limitada de tareas concurrentes.

En un primer contacto con el Big Data, hay quien se plantea si no es *Business Intelligence* o inteligencia empresarial con otro nombre. Ante esto hay que tener en cuenta que, si bien en ambos casos se maneja un volumen importante de datos para ayudar en el análisis de la información y la toma de decisiones, en el Big Data se integra información de una mayor diversidad de fuentes (internas y externas) y formatos (variedad), y en muchos casos el resultado se ha de obtener con mucha mayor celeridad (velocidad).

Por otra parte, desde un punto de vista de arquitectura y tecnología, se puede estructurar un sistema Big Data en cinco capas principales:

• Capa de fuentes de datos: en esta capa estarían todos los orígenes de la información, desde bases de datos relacionales hasta cualquier tipo de datos, estructurados o no.

• Capa de integración: aquí se adquieren los datos y se integran en conjuntos con el formato adecuado.

• Capa de almacenamiento de datos: el conjunto de recursos adecuados para el almacenamiento de grandes volúmenes de datos.

• Capa de análisis y modelos de computación: esto incluye diversas herramientas de manejo de datos, que operan sobre los recursos de almacenamiento e incluyen la gestión de los datos y los modelos de programación.

• Capa de presentación y aplicación: incluye las tecnologías de visualización tales como dispositivos móviles, navegadores, etc. Una vez obtenido el conocimiento, éste se puede aplicar en distintos procesos.

Dentro de las tecnologías utilizadas destacan las herramientas open source[2], que integran el almacenamiento y procesado de datos, la gestión del sistema y otros módulos para ofrecer una solución completa.

I.2.- RIESGOS LEGALES, AMENAZAS Y OPORTUNIDADES.

Como acabamos de comentar, los tratamientos de Big Data analizan grandes volúmenes de datos a una velocidad antes inimaginable. Por si ello fuera poco, se han revelado como una herramienta eficaz y útil para realizar predicciones. Es indudable su valor en sectores clave como en el sanitario, donde existen ya muchos ejemplos de su eficacia para reducir el tiempo de ingreso hospitalario o predecir futuras enfermedades y riesgos sanitarios.

También se prevé su utilización en las *Smart Cities* como herramienta para prevenir, por ejemplo, colapsos de tráfico y excesos de contaminación. En el sector de la distribución permite anticiparse al consumidor evitando situaciones de desabastecimiento de productos y falta de suministro. Por estos y otros usos no planteados, se puede considerar que esta tecnología será muy provechosa para la sociedad ya que puede aportarle numerosos y valiosos beneficios económicos y sociales.

Pero como toda herramienta potente y novedosa, surgen dudas y preocupaciones sobre posibles usos que, o bien no sean lícitos por realizarse sin respaldo legal para ello, o bien generen abusos en usos basados en el valor económico de los datos personales, considerados como el petróleo del siglo XXI.

La generación de perfiles de consumidores o profiling es sin duda uno de los usos principales del Big Data, y puede entrañar riesgos por posibles tratamientos basados en predicciones, si se utilizan de forma discriminatoria excluyendo a sectores minoritarios en base a los resultados analizados, lo que se viene denominando "la dictadura de los datos". También surgen temores fundados sobre su potencial uso en sectores poblacionales vulnerables como pueden ser menores, ancianos o colectivos marginados, por lo que es necesario establecer garantías adecuadas en todos los ámbitos.

2 "*Open Source*" o código abierto es una modalidad colaborativa y pública de desarrollo de programas cuyo código fuente se distribuye con una licencia que permite a cualquiera y de forma gratuita estudiarlo, modificarlo y redistribuirlo, requiriendo generalmente que los cambios sean a su vez publicados y limitado o prohibiendo cobrar por los mismos.

Basándose en los riesgos que originan estos tratamientos para la privacidad de las personas, se han identificado en este "Código de buenas prácticas en protección de datos para proyectos de Big Data", los aspectos que deben ser abordados para que los tratamientos de Big Data sean conformes a la normativa sobre protección de datos personales.

Un primer bloque lo componen los aspectos legales. Abordan cuestiones clave, como la trasparencia que se debe ofrecer en la información previa facilitada a los afectados, así como la obtención del consentimiento y el ejercicio de derechos por parte de los afectados.

No debemos olvidar la dificultad práctica que nos podemos encontrar con futuros usos no previstos en el momento de obtener la información y el consentimiento para ellos, así como las peculiaridades legales derivadas de la generación de perfiles y la monitorización de la conducta (profiling), del ejercicio del derecho de oposición y el derecho de impugnación de valoraciones basadas en decisiones automatizadas.

También se encuentran dificultades añadidas debido al origen y procedencia de la información, que no siempre proviene de fuentes propias sino también de terceros, y al uso de los datos por diferentes figuras (responsables y encargados de tratamiento). Finalmente se plantean cuestiones diversas como el plazo de conservación y retención de los datos, así como el uso y la reutilización de datos previamente disociados.

Un segundo bloque de riesgos y amenazas viene de la mano de aspectos técnicos y de seguridad. Así, la anonimización, debe ser considerada teniendo en cuenta el riesgo de reidentificación de la información, lo que obliga a realizar un análisis del riesgo para eliminarlo o minimizarlo, e incluso una evaluación de impacto. También plantean peculiaridades el uso del cloud computing en los tratamientos de Big Data, que obliga a plantearse estrategias de seguridad reforzadas, como el uso de técnicas de cifrado.

I.3.- ÉTICA DIGITAL, PRIVACIDAD Y BIG DATA.

El uso del Big Data, como la mayoría de las tecnologías con incidencia directa en las personas, y en definitiva en la sociedad, no está exento de preocupaciones o, desde una perspectiva menos severa, de ciertas consideraciones previas que hay que tener en cuenta. ¿Hasta dónde queremos y podemos llegar como "sociedad tecnológica"? ¿Los beneficios que nos aportará esta nueva forma de explotación de la información están justificados a cualquier precio? ¿Debemos plantearnos aspectos éticos sobre el uso que se le dará a estas ingentes cantidades de información y sobre las decisiones que en base a esa información se puedan tomar sobre cientos de miles de individuos?

La respuesta parece clara, y es que al menos debemos plantearnos tales consideraciones. Ya lo han hecho autoridades de control en materia de privacidad, como el Supervisor Europeo de Protección de Datos[3], que ha puesto el énfasis en los siguientes desafíos que puede traernos (en realidad, que ya nos ha traído) el Big Data:

• Falta de transparencia, ligada al celo que las organizaciones ponen cada vez más en cómo procesan la información y qué utilidad puede tener el resultado de ese procesamiento para la propia organización. Este celo en no revelar los tratamientos de la información que llevan a cabo puede conducir a que los ciudadanos no sepan realmente qué ocurre con sus datos una vez los facilitan. Y esto es así porque, en ocasiones, quizás ni las propias empresas son del todo conscientes de hasta dónde llegarán tales tratamientos de información. En este sentido, es muy relevante que los interesados sean conocedores de los impactos que los diferentes tratamientos pueden tener sobre su privacidad.

• Desequilibrio en la información entre las personas y las empresas que tratan sus datos personales; desequilibrio que es muy probable que aumente con el avance de sistemas Big Data. Un ejemplo relevante de esta posible situación es la modificación de precios de un producto en función de lo que esté dispuesto a pagar cada consumidor, según los resultados que haya arrojado el correspondiente análisis de la información a disposición de una empresa[4].

3 Supervisor Europeo de Protección de Datos / Opinion 7/2015 Meeting the challenges of Big Data / 19 de noviembre de 2015.
4 Iniciativas como el http://www.datatransparencylab.org/ trabajan en proponer iniciativas para mejorar la transparencia en la provisión de servicios de la Sociedad de la Información.

Resulta evidente que los dos anteriores aspectos se cruzan de forma directa con los principios éticos y de privacidad básicos. El derecho a la protección de datos, intrínsecamente relacionado con el derecho a la intimidad, está basado en los principios relativos al tratamiento, recogidos en el artículo 5.1 del RGPD donde se pone de manifiesto que los datos personales sólo podrán ser tratados de manera lícita, leal y transparente. Además, los datos deberán estar limitados a lo necesario en relación con fines determinados, explícitos y legítimos.

Otro principio fundamental del derecho a la Protección de Datos es el de legitimación, directamente relacionado con el deber de información, por el cual el titular de los datos debe ser informado con absoluta claridad acerca de las finalidades para las que se recaban sus datos. El artículo 13 del RGPD incluye toda la información que se debe facilitar a los interesados cuando se recaban sus datos personales. El Big Data, por su propia esencia, puede llevar a situaciones en los que la finalidad inicial para la que se recogió el dato quede al menos "difuminada" una vez el dato es explotado.

En conclusión, la regulación tendrá un papel determinante en la contención de los tratamientos de información de manera que se ajusten a los principios éticos y de privacidad que imperan en cada vez más áreas del mundo. Y junto a esa regulación, sin duda las autoridades regulatorias y de control tendrán también mucho que decir.

I.4.- MARCO DE LA GOBERNANZA.

El contexto de la gobernanza de la protección de los datos personales, entendida como el conjunto de metodologías, políticas y herramientas que permiten la gestión de los datos personales y de la información para asegurar su calidad, su control y explotación según los objetivos estratégicos definidos dentro de una empresa u organización y el cumplimiento de la normativa sobre esta materia, ha sufrido un cambio radical con el desarrollo de las tecnologías Big Data.

El impacto inicial se ha producido por la gran cantidad de información generada y que tiene que ser almacenada. Además, se une a ese gran volumen la naturaleza heterogénea de su contenido. Se manejan datos de tipo estructurado tradicionales, pero también información de toda la actividad producida por los usuarios, como audio, video, imágenes, conversaciones, que son muy difíciles de tratar con las herramientas que existían anteriormente.

La rapidez con la que se tienen que almacenar unida a esa extensión en su tipología, hace muy difícil que los procesos de verificación y calidad utilizados hasta ahora sean totalmente eficaces, por lo que es necesario crear nuevas metodologías y herramientas adecuadas.

Y para esto, es fundamental que toda la organización se involucre en la obtención, análisis y comprensión de los datos y de la información disponible. Estas tareas no pueden quedarse sólo en las áreas de sistemas de información y de inteligencia de negocio. Todos los miembros de la organización deben ver el dato como un valor en sí mismo, y tienen que tener en cuenta, desde los primeros momentos de la definición de un producto, servicio o proceso, cómo conseguir la información adecuada, cómo almacenarla, cómo usarla para mejorar el propio servicio o proceso, y cómo analizarla posteriormente.

Cada persona dentro de la organización deberá participar en la comprensión de la información, desde el perfil más técnico al más ejecutivo. Para conseguir la máxima eficacia y valor del dato, se tienen que crear canales de comunicación y un lenguaje común dentro de la organización para forzar y gestionar adecuadamente esta dinámica.

Pero la extensión del valor de los datos a toda la organización conlleva un riesgo en la gestión de información confidencial y personal. El reto principal es cómo conseguir que la información más valiosa sea utilizada por las personas más adecuadas sin comprometer la privacidad y confidencialidad de los datos.

Se deben implantar los sistemas de control de acceso, monitorización y anonimización necesarios, para que el análisis se adecúe a las normativas de protección de datos, unido a la creación de políticas preventivas y mitigadoras de riesgos que puedan cubrir aspectos no contemplados por esos mismos marcos legales.

Y asociado a este último punto, destaca la gestión del consentimiento de los usuarios y la transparencia sobre cómo se utiliza su información personal. El usuario debe conocer en todo momento, de forma sencilla, qué información personal se utiliza y para qué se utiliza, así como permitir que pueda o no dar su consentimiento, e incluso posteriormente oponerse al tratamiento.

I.5.- PRÁCTICAS HABITUALES EN LOS TRATAMIENTOS.

Las organizaciones en sus tratamientos de Big Data tratan de obtener valor, ya sea económico en las organizaciones con ánimo de lucro u ofrecer un mejor servicio en las que no lo tienen.

Aunque el término se ha puesto de moda en los últimos años, se lleva haciendo desde hace bastante tiempo y se pueden encontrar ejemplos relevantes en el pasado.

Un ejemplo interesante es cómo Wal-Mart, la gran cadena de supermercados de bajo coste de Estados Unidos, distribuyó con anticipación lo que iban a comprar los ciudadanos al acercarse el huracán Katrina. Esta preparación, en base al análisis de qué compraban sus clientes en estos eventos, le permitió estar preparado y poder satisfacer ese pico de demanda atípica.

Los casos de uso y negocio se pueden agrupar en seis categorías:

- Mejor conocimiento del cliente: la información permite ofrecer un mejor servicio y atención al cliente.

- Mejor conocimiento del mercado para la captación de nuevos clientes.

- Personalización de productos y servicios: la información permite personalizar el servicio ofreciendo una mejor experiencia de cliente, incrementando la fidelización y satisfacción.

- Mejora y rapidez en la toma de decisiones: la información permite a las organizaciones públicas y privadas tomar mejores decisiones, optimizando la gestión de procesos y, por tanto, reduciendo costes aumentando la competitividad.

- Previsión del comportamiento: un análisis adecuado permite obtener una mejor visión de qué puede pasar, ampliar la visión estratégica y de negocio, crear nuevos servicios y productos, y obtener nuevos ingresos.

- Monetización: la propia información puede ser monetizada, por ejemplo, a través de una mejor publicidad o compartiendo estos datos con otras compañías (eso sí, asegurando el cumplimiento del marco legal).

Prácticamente en todos los sectores de actividad se pueden encontrar ejemplos de uso del Big Data. Aunque solo se llegan a conocer los casos de éxito, hay que estar preparado para que en cualquier uso de Big Data se produzcan resultados erróneos y poder evitar que su aparición cause efectos dañinos. En muchos casos, esos “errores” pueden ser inocuos (pensemos, por ejemplo, en la presentación de una publicidad ligeramente errónea donde el perjuicio es que la probabilidad de venta es cero) comparado con el uso para la toma de decisiones que puedan causar perjuicios a personas o colectivos específicos.

Para apreciar los usos del Big Data, nada más útil que presentar algunos ejemplos reales en varios sectores de actividad:

• Venta por Internet: es el más obvio y uno de los pioneros. Todo aquel que compra por Internet puede observar cómo la publicidad que recibe es cada vez más atinada al contenido de las búsquedas que realiza o la información que consulta, incluso aunque aparentemente no se haya identificado. Asimismo, en el proceso de compra, rara es la tienda que no ofrecerá productos y servicios complementarios en base a la experiencia de otros clientes.

• Venta presencial: análisis de los patrones de compra dentro de la tienda y por cliente. En el caso del cliente, el uso de las tarjetas de fidelización es clave. Toda esa información ayuda a colocar los productos para maximizar su venta, ofrecer descuentos y productos a los clientes apropiados, identificar compras correlacionadas, etc. Cuanto más conozcan cómo compramos, más se puede adaptar la tienda para maximizar su venta.

• Sector bancario: uso en el análisis de riesgos en general y en la concesión de préstamos en particular, lucha contra el fraude, personalización de ofertas para clientes, captación de clientes externos ayudándoles a utilizar su información financiera o a crear servicios de valor añadido, etc.

• Industria petrolífera: ha sido una de las industrias pioneras en el uso del Big Data, el apropiado análisis de la información sísmica y otros datos geológicos permite perforar en los lugares más productivos. Una vez en producción, la monitorización continua mediante múltiples sensores permite maximizar el tiempo de funcionamiento y mejorar la seguridad de trabajadores y del entorno.

• Políticas públicas: Soporte para toma de decisiones en el desarrollo de políticas públicas en diferentes ámbitos como el educativo, sanitario, servicios de emergencia, turismo, transporte, seguridad ciudadana, empleo.

• Ayuda al desarrollo y situaciones de emergencia: Herramienta para la gestión de situaciones de catástrofes, políticas de desarrollo humano y social.

II. NORMAS Y PRINCIPALES OBLIGACIONES LEGALES EN MATERIA DE PRIVACIDAD.

II.1.- RÉGIMEN JURÍDICO APLICABLE.

El marco jurídico aplicable a los proyectos de Big Data se compone por un conjunto de normas que lo regulan, aunque ninguna es específica para este tipo de tratamiento. Aunque actualmente la legislación a aplicar es la Ley 15/1999 de protección de datos de carácter personal (en adelante, LOPD) y su reglamento de desarrollo, el Reglamento General de Protección de Datos (en adelante, RGPD) aprobado el 18 de abril de 2016, será plenamente aplicable a partir del 25 de mayo de 2018, por lo que debe ser tenido en cuenta por cualquier entidad que pretenda desarrollar proyectos con este tipo de tecnología.

El RGPD se aplica al tratamiento de datos personales, entendiéndose estos como cualquier información concerniente a personas físicas identificadas o identificables. La identificabilidad, que supone la aplicación de la normativa, se refiere a que una persona pueda ser identificada por un dato o por la combinación de información de diversas fuentes. De manera más precisa, para determinar si una persona es identificable, han de usarse "todos los medios que puedan ser razonablemente utilizados y sin esfuerzos desproporcionados". El análisis de la identificabilidad debe basarse en dos criterios, el de la razonabilidad en la disponibilidad de los medios (técnicos, humanos y fuentes de datos) y en la proporcionalidad de los esfuerzos para poder identificar directa o indirectamente a la persona física por parte del responsable del tratamiento o por cualquier otra persona.

En ese sentido, cuando no sea posible la identificación de los individuos, o esta requiera esfuerzos desproporcionados, no será de aplicación la normativa de protección de datos. Además, hay muchos casos en los que los datos tratados y analizados con Big Data no serán datos personales (por ejemplo, aquellos datos relativos al tráfico, la contaminación o al clima), por lo que tampoco será aplicable el RGPD.

Además, si los datos están completamente anonimizados, no se considerarán datos de carácter personal, por lo que la normativa tampoco será de aplicación. En este sentido, la anonimización supone que no será posible identificar a la persona con datos o con información de diversas fuentes, teniendo en cuenta todos los medios que puedan ser razonablemente utilizados para su identificación. El Grupo de Trabajo del Artículo 29 (GT29) ha analizado este concepto de dato personal en su Dictamen 04/2007[5].

Pero en muchos otros casos, como se analiza a continuación, sí existirá tratamiento de información personal (medios sociales, transacciones bancarias, etc). Por tanto, siempre que sea posible identificar a los individuos, habrá de tenerse en cuenta la aplicación de la normativa de protección de datos. Y para ello, será necesario contar con alguno de los fundamentos legales que establece la normativa.

Al respecto, junto con el consentimiento del interesado para el tratamiento de sus datos personales, el artículo 6 del RGPD también establece otras legitimaciones, como el tratamiento de datos necesarios para la ejecución de un contrato, cuando haga referencia al cumplimiento de una obligación legal, cuando el tratamiento de los datos tenga por finalidad proteger un interés vital del interesado, o cuando el tratamiento sea necesario para el cumplimiento de una misión realizada en interés público o para la satisfacción del interés legítimo perseguido por el responsable del tratamiento o por un tercero, siempre que no se vulneren los derechos y libertades fundamentales del interesado.

En el caso que nos ocupa, para el tratamiento de datos en proyectos de Big Data, se analizarán los fundamentos más relevantes. Respecto a la primera de las legitimaciones para el tratamiento, el consentimiento, este deberá reunir las condiciones que establece la normativa vigente.

5 Dictamen 4/2007 sobre el concepto de datos personales (WP 136).

Es importante resaltar que, para que el consentimiento cumpla con los requisitos previamente mencionados, será necesario que éste sea libre, explícito y que pueda ser revocado. El artículo 7 del RGPD regula específicamente las condiciones del mismo.

No obstante, otras legitimaciones pueden tener cabida para el tratamiento de datos personales. Por ejemplo, en caso de que estos datos sean necesarios para el desarrollo de contrato o precontrato de una relación negocial, laboral o administrativa, entre el afectado y el responsable, y sean necesarios para su mantenimiento o cumplimiento. En estos casos, habrá que determinar cuándo el tratamiento es necesario para el cumplimiento de este contrato, dado que normalmente, los proyectos de Big Data suelen tener un alcance mayor al inicio, por ejemplo, de la prestación de un servicio, mediante el análisis de los datos resultantes de ese servicio para otras finalidades. En ese sentido, la utilización de este fundamento habrá de ser cuidadosamente analizado caso por caso.

Es también posible que los tratamientos de Big Data puedan estar basados en la satisfacción de un interés público o en el ejercicio de poderes públicos. Normalmente, este tipo de tratamientos será llevado a cabo por entidades públicas, aunque sería posible identificar casos en que la atención de intereses públicos relevantes pudiera asumirse por entidades privadas.

Ejemplos de tratamientos de Big Data, con o sin el empleo de datos personales, sobre esta base jurídica podrían ser los relacionados con proyectos de “Smart Cities” o los desarrollados por servicios públicos de salud.

En todo caso, el Reglamento General de Protección de Datos precisa que tanto el interés público como los poderes públicos que se ejerzan tienen que estar establecidos en el Derecho de la Unión o de los Estados Miembros. Asimismo, el Reglamento prevé, y ello es relevante en el contexto de tratamientos de Big Data, que para determinadas finalidades, como pueden ser las relacionadas con un interés público esencial, las de investigación científica, las relacionadas con la atención sanitaria o social, o las relativas a salud pública, los tratamientos serán posibles en las condiciones que determine la legislación europea o nacional. Las normas correspondientes establecerán, además, las garantías necesarias para la protección de los derechos y libertades de los interesados.

Otra de las posibilidades que ofrece la normativa es la que se centra en la satisfacción del interés legítimo perseguido por el responsable del tratamiento, siempre que no se vulneren los derechos y libertades fundamentales del interesado. En estas situaciones será necesario analizar el balance entre el interés legítimo y los derechos y libertades. El Grupo de Trabajo del Artículo 29 se ha pronunciado al respecto en su opinión Dictamen 06/2014 [6].

Dicho Dictamen establece unos criterios en el balance del interés legítimo del tratamiento que deberán ser objeto de reflexión en los análisis de impacto que en su caso lleve a cabo el responsable del tratamiento:

-Que exista un interés legítimo del responsable o del tercero que alegue dicho interés.
-El impacto que dicho tratamiento tenga en el interesado.
-La naturaleza de los datos objeto de tratamiento y la forma de dicho tratamiento.
-Las expectativas razonables de los interesados en relación con el tratamiento.
-El desequilibrio entre el responsable del tratamiento y el interesado.

6 Dictamen 06/2014 sobre el concepto de interés legítimo del responsable del tratamiento de los datos en virtud del artículo 7 de la Directiva 95/46/CE (WP 217).

Una de las cuestiones básicas y previas es el primer elemento del análisis, la existencia de un interés legítimo del responsable. El Dictamen menciona varios casos en los que dicho interés puede existir, tales como la libertad de información y expresión, las actividades de marketing o publicidad, prevención del fraude o mal uso de servicios, seguridad, finalidades científicas, estadísticas o de investigación.

El Dictamen hace también mención a la personalización de ofertas comerciales y actividades de marketing online y offline. No obstante, advierte, que aun existiendo ese interés legítimo, éste no es base suficiente para la ejecución de complejos perfilados de clientes que representarían una intrusión significativa en su privacidad. Estaríamos en este caso en un impacto sobre los interesados que hay que tener en cuenta en el mencionado análisis de balance. Por tanto, la existencia de un interés legítimo no es base suficiente, pero sí necesario en el análisis. Habrá que tener en cuenta el impacto del tratamiento en los derechos fundamentales y libertades de los interesados.

Algunas de las aplicaciones de Big Data tienen una finalidad estadística, cuyo objetivo es obtener datos estadísticos agregados, que en muchos casos permitan tomar decisiones públicas o de negocio. Así, el RGPD define la finalidad estadística como cualquier operación de recogida y tratamiento de datos personales necesarios para la producción de resultados estadísticos. Esta finalidad implica que el resultado del tratamiento con fines estadísticos no sean datos personales, sino datos agregados, y que este resultado no se utilice para respaldar medidas o decisiones relativas a personas físicas concretas. El RGPD establece que los fines estadísticos no se considerarán incompatibles con los fines iniciales, si bien el Reglamento menciona que el responsable debe incluir garantías adecuadas en el tratamiento que aseguren que se aplican medidas técnicas y organizativas para garantizar que no se puede identificar a los interesados.

Lo mismo cabe decir de otras finalidades relacionadas con Big Data tales como la científica o de innovación donde se ofrece una regulación favorable en estos mismos términos.

Además, es necesario tener en cuenta que la normativa dota de una serie de garantías a los titulares para la protección de sus datos, entre los que figura el principio de finalidad, que debe ser determinada, explícita y legítima, para la que se hayan obtenido los datos. Este principio de finalidad puede suponer una de las barreras para los proyectos de Big Data, dado que no siempre se conoce desde el comienzo el alcance del proyecto.

En ese sentido, es necesario hacer referencia a que los datos personales no podrán usarse para finalidades incompatibles con aquellas para las que los datos hubieran sido recogidos, lo que no significa que no puedan utilizarse para finalidades diferentes para las que se recogieron, si no que éstas no deben ser incompatibles. El Grupo de Trabajo del Artículo 29 describe de forma pormenorizada en su Dictamen sobre el principio de finalidad antes mencionado en qué casos nos encontramos ante fines incompatibles.

El análisis de no incompatibilidad es básico en el Big Data, dado que, en buena medida, basa sus analíticas en el tratamiento posterior con finalidades adicionales a la finalidad original. El Grupo de Trabajo del Artículo 29 ha analizado este aspecto en su Dictamen (WP 203)[7]. Al respecto, para saber si los usos posteriores de los datos personales son compatibles, el Dictamen establece los siguientes criterios:

-Debe existir una relación entre la finalidad original y la finalidad o finalidades ulteriores.
-El tratamiento ulterior debe encontrarse dentro de las expectativas razonables del interesado.
-Debe tenerse en cuenta la naturaleza de los datos objeto de tratamiento y la sensibilidad de los mismos.
-Debe considerarse el impacto que este tratamiento va a tener en los interesados.
-Deben considerarse las medidas de protección que el responsable del tratamiento establece, en particular las medidas técnicas y organizativas: encriptación, seudonimización, separación funcional, transparencia, oposición al tratamiento.

7 Dictamen sobre el principio de finalidad (WP 203)

El artículo 6.4 del RGPD ha incorporado a la norma estos criterios, haciendo mención expresa al cifrado y la seudonimización dentro de las garantías adecuadas, con lo que serán de aplicación directa al análisis de compatibilidad de todos aquellos tratamientos posteriores que no estén basados en el consentimiento del interesado y, por lo tanto, deberán tenerse en cuenta en la evaluación de impacto que realice el responsable sobre dichos tratamientos.

Otros principios a tener en cuenta son los principios de minimización y conservación de los datos. La normativa establece que los datos sólo se podrán recoger para su tratamiento, así como someterlos a dicho tratamiento, cuando sean adecuados, pertinentes y limitados en relación con fines determinados, explícitos y legítimos. El principio de minimización constituye uno de los elementos más importantes de la normativa de protección de datos aplicables a este entorno. Este principio va intrínsecamente unido al de conservación de los datos por el tiempo que sea necesario para la finalidad para la que hubieran sido recogidos.

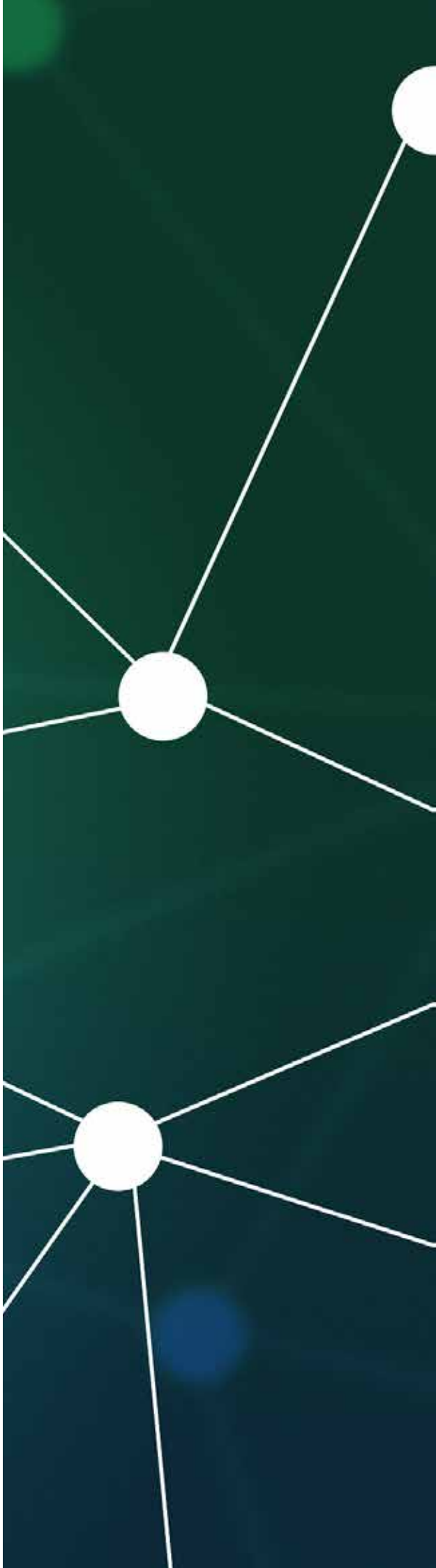
En el entorno Big Data este principio es especialmente relevante, dado que las diferentes fuentes a la que se tiene acceso pueden dar lugar a una recogida masiva de datos que no sean pertinentes para la finalidad del tratamiento. Las organizaciones deberán tener en cuenta que no se recojan datos excesivos en relación a esa finalidad. Tampoco deberían conservarse por más tiempo del que sea necesario, periodo tras el cual deberían eliminarse.

Por otro lado, la posibilidad de que en un futuro se puedan llevar a cabo tratamientos de Big Data con diversas finalidades no significa, en ningún caso, que se recojan, "por si acaso", más datos de las personas que aquellos que son necesarios para la finalidad primaria que motiva su recogida.

Asimismo, hay que mencionar el principio de la seguridad de la información. La normativa establece que se deberán adoptar las medidas técnicas y organizativas apropiadas para garantizar un nivel de seguridad adecuado al riesgo. Para evaluar dicha adecuación, se tendrán en cuenta en particular los riesgos asociados a destrucción, pérdida o alteración, o a la comunicación o acceso no autorizados.

Al respecto, la Agencia Europea de Seguridad de la Información y la Red (ENISA), en su informe Big Data Security ha identificado las diferentes amenazas informáticas actuales y emergentes que puede haber para los proyectos de Big Data. Especial cuidado se tendrá en relación con los datos que estén alojados en la nube, aspectos sobre los que conviene revisar el Dictamen 05/2012[8] del Grupo de Trabajo del Artículo 29 en el que destaca que aquellas empresas que vayan a utilizar servicios en la nube, tendrán que tener en cuenta la seguridad de los proveedores.

Respecto a la necesidad de respetar los derechos de acceso, rectificación, cancelación y oposición que tienen los afectados, hay que afirmar rotundamente que también pueden ejercerse en el tratamiento de datos en proyectos de Big Data en los que se traten datos de carácter personal. Asimismo, tendrá derecho a que el responsable limite el tratamiento de sus datos y a no ser objeto de una decisión basada únicamente en tratamientos automatizados.



8 Dictamen 05/2012 sobre la computación en nube (WP 196)

II.2.- RESPONSABLE Y ENCARGADO DEL TRATAMIENTO.

El concepto de responsable y encargado del tratamiento es crucial en el contexto del tratamiento de datos y lo mismo sucede en los proyectos de Big Data, ya que con frecuencia se suelen externalizar los tratamientos y el análisis de los datos.

La legislación establece normas específicas para aquellos casos en los que los datos son tratados por diferentes actores. En ese sentido, el RGPD establece que el responsable, es la "persona física o jurídica, autoridad pública, servicio u otro organismo, que determine los fines y medios del tratamiento, y asimismo define encargado como persona física o jurídica autoridad pública, servicio u otro organismo que trate datos personales por cuenta del responsable del tratamiento".

El responsable del tratamiento fija la finalidad del tratamiento y decide sobre la externalización del mismo y en qué grado delega las actividades de tratamiento a otra organización. Además, elegirá únicamente a un encargado que ofrezca garantías suficientes para aplicar medidas técnicas y organizativas adecuadas de manera que el tratamiento sea conforme al RGPD.

Además, el RGPD introduce la figura del corresponsable del tratamiento en los casos en los que dos o más responsables determinen conjuntamente los objetivos y los medios del tratamiento, que han de fijar de mutuo acuerdo sus responsabilidades respectivas en el cumplimiento de las obligaciones impuestas por el RGPD, debiéndose poner a disposición del interesado los aspectos esenciales del acuerdo.

Por otro lado, cuando un tercero, actuando en nombre del responsable, suministra los medios o la plataforma en el caso de cloud computing se considera que es el encargado. Otros casos pueden también darse en la instalación y mantenimiento de herramientas informáticas de Big Data, contratación de empresas para que realicen o analicen el Big Data, etc.

Es importante analizar y precisar el rol de cada uno de estos actores para determinar sus obligaciones en relación con la legislación sobre protección de datos. El Dictamen 1/2010[9], establece que, para poder actuar como encargado del tratamiento, tienen que darse dos circunstancias: que se trate de una entidad independiente del responsable y, segunda, que se traten los datos por cuenta de éste. Además, también puede llevar a cabo actividades específicas sobre el tratamiento, con autonomía para determinar qué medios técnicos son los más adecuados.

No obstante, si el encargado del tratamiento recoge, trata datos por su cuenta o establecen un nuevo vínculo jurídico con los titulares de los datos suministrados por

el responsable, determinando la finalidad y tratándolos conforme a la misma, se le considerará responsable respecto a este nuevo tratamiento.

El RGPD establece que la realización de tratamientos por cuenta de terceros deberá estar regulada en un contrato por escrito, u otro acto jurídico, que vincule al encargado respecto del responsable y establezca el objeto, la duración, la naturaleza y la finalidad del tratamiento, así como el tipo de datos personales, categorías de interesados y las obligaciones y derechos del responsable.

Habrá ocasiones en que el encargado del tratamiento necesite recurrir a terceros, como cuando se contrate servicios en la nube, que pueda suponer la participación de terceros que actúen como subencargados del tratamiento.

En estos casos, el encargado que subcontrate servicios deberá comunicarlo al responsable, informando sobre el tipo de servicio que se ha subcontratado, así como las garantías que estas organizaciones ofrecen para cumplir con la normativa. En cuanto a las obligaciones de los subcontratados, estas serán las mismas que las que se aplican a los encargados, y han de recogerse asimismo en un contrato.

La ley establece que una vez cumplida la prestación contractual, los datos de carácter personal deberán ser destruidos o devueltos al responsable del tratamiento, al igual que cualquier soporte o documentos en que conste algún dato de carácter personal objeto del tratamiento.

Al igual que en otros tratamientos, en aquellos casos en que el encargado del tratamiento destine los datos personales a otra finalidad, los comunique o los utilice incumpliendo las estipulaciones del contrato, será considerado también responsable del tratamiento y, por tanto, deberá informar al interesado sobre la identidad del nuevo responsable y de las categorías de datos que se van a tratar, así como de los destinatarios de los datos.

En cada caso habrá que estudiar la asignación de responsabilidad de tal manera que el cumplimiento de las normas de protección de datos se vea garantizada.

9 El Dictamen 1/2010 establece que «el papel primero y primordial del concepto de responsable del tratamiento es determinar quién debe asumir la responsabilidad del cumplimiento de las normas sobre protección de datos y de qué manera los interesados pueden ejercer sus derechos en la práctica. En otras palabras, debe asignar la responsabilidad».

II.3.- PRINCIPALES IMPLICACIONES DE LOS TRATAMIENTOS BIG DATA EN PRIVACIDAD.

II.3.1.- ORIGEN DE LOS DATOS.

El origen de datos es el primer aspecto que debe tenerse en cuenta en la cadena de tratamientos contemplados en un sistema de Big Data. Una parte importante de la complejidad del análisis de estos tratamientos ocurrirá en aquellos casos en que el sistema se nutra de información proveniente de múltiples orígenes.

Según el nivel de confiabilidad que ofrezcan los diferentes orígenes de datos, la calidad de los datos primarios puede quedar comprometida de inicio y arrastrarse durante todo su ciclo de vida.

Aquí es donde la clasificación de las fuentes en endógenas (circunscritas a la definición y control de la propia organización) y exógenas (fuentes externas) provocará la necesidad de aplicar filtros y controles compensatorios en mayor o menor medida.

Más allá de la calidad de los datos primarios, la integración de los mismos desde diferentes orígenes, o fuentes de procedencia, no siempre es sencilla pese a la aplicación de sofisticadas técnicas de depuración apoyadas a menudo en diccionarios de datos. Esto hace que deba recurrirse a otros datos complementarios para ayudar a garantizar la fiabilidad de los datos incorporados.

Para minimizar estas dificultades, cada vez se obtienen en origen un mayor número de metadatos (información secundaria constituida por datos que califican a otros datos) que pueden informar de la ubicación desde donde se ha recabado el dato, en qué momento, quién lo ha facilitado, desde qué tipo de sensor, su factor de precisión, etc.

Es fácil intuir que la adición indiscriminada de metadatos redunda en ventajas e inconvenientes, especialmente para la privacidad:

- Como ventajas, aporta una mejor interoperabilidad y aumento de la calidad de los datos personales tratados (o datos visibles), al estar calificados por los metadatos (o datos invisibles).

- Como inconvenientes, representa elevar el nivel de riesgo de incumplir algún principio general de la protección de datos, como es el principio de minimización de los datos, y el principio de limitación de la finalidad al posibilitar que los metadatos sean tratados con finalidades incompatibles respecto a las previstas para la información primaria.

A partir de estas fortalezas y debilidades, deberá buscarse el equilibrio en cada caso particular. Así surge con fuerza el concepto de Protección de Datos desde el Diseño y por defecto (PDdD – o PbD en inglés), señalado en el art. 25 RGPD, apoyándose en una Evaluación de Impacto relativa a la Protección de Datos (EIPD), según dispone el art. 35 RGPD.

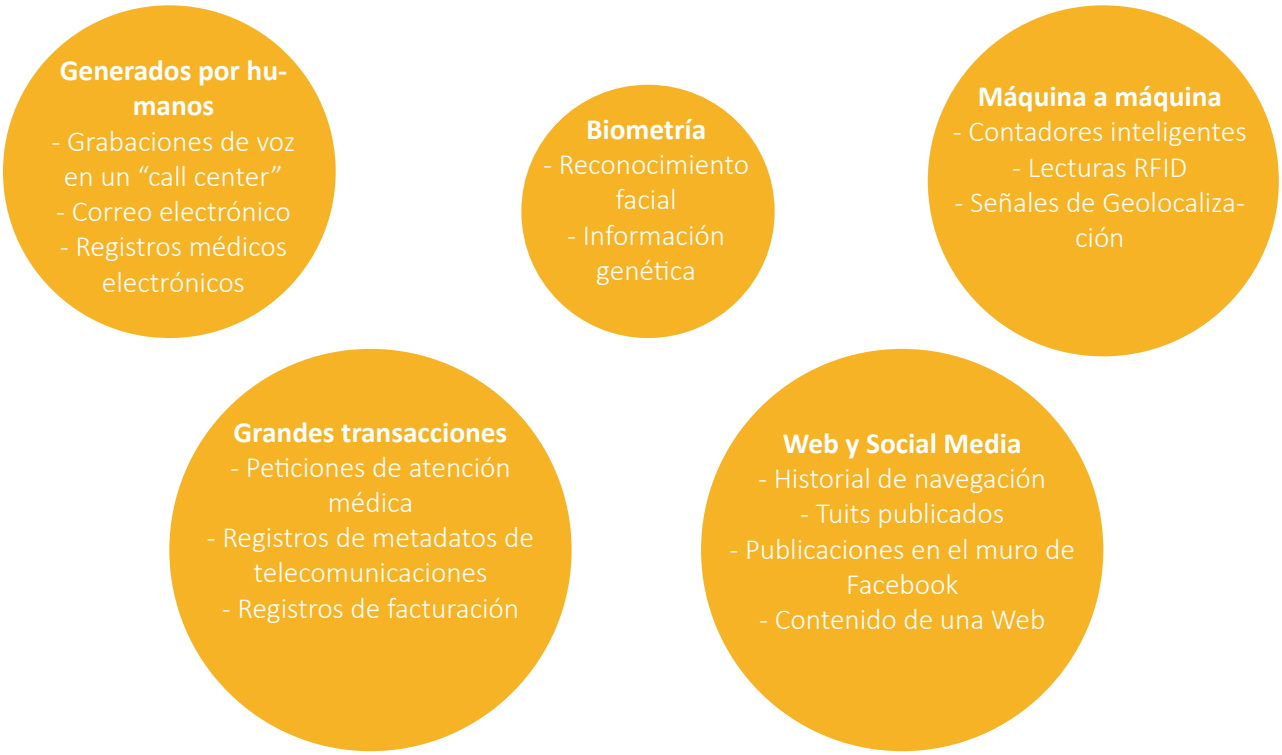
Para ilustrar la enorme desproporción, según el caso, entre los datos primarios o visibles y los metadatos, podemos atender a algunos ejemplos significativos:

- Un tuit. Los datos primarios ocupan 140 caracteres, mientras que los más de 30 metadatos asociados pueden llegar a superar en varias veces esa cifra.

- Una fotografía con el móvil. Acompañan por defecto más de 40 metadatos, incluyendo el modelo de móvil, las coordenadas GPS y la altitud en el momento de capturar la imagen.

Llegados a este punto, cabe recordar que la legislación vigente en materia de protección de datos, incluyendo el principio de consentimiento, aplica a todos los datos personales con independencia de que éstos sean primarios o metadatos (visibles o invisibles).

A modo de ejemplo, Sunil Soares en 2012 hizo una clasificación bajo el concepto de los orígenes de datos bajo los conceptos:



II.3.2.-TRANSPARENCIA EN LA INFORMACIÓN.

En el análisis de la interrelación entre consentimiento, y derecho a la información en relación con la técnica del Big Data, en la 36ª Conferencia Internacional de Autoridades de Protección de Datos y Privacidad, se emitió una Resolución sobre Big Data, en la que se señala que la protección proporcionada por los principios de información y consentimiento es más importante que nunca, sobre todo en momentos en que se recopila una cantidad cada vez mayor de información sobre nosotros.

El RGPD hace una apuesta importante por el hecho de que cada ciudadano tenga un mayor poder de disposición sobre sus datos de carácter personal. Por otro lado, la realización de las técnicas Big Data y la libre circulación de los datos constituyen una realidad innegable para las empresas y, por tanto, esta nueva realidad jurídica no debe impedir el desarrollo económico de la zona UE. Así lo señala el Considerando 13 del nuevo Reglamento, cuando afirma que: "el buen funcionamiento del mercado interior exige que la libre circulación de los datos personales en la Unión no sea restringida ni prohibida por motivos relacionados con la protección de las personas físicas en lo que respecta al tratamiento de datos personales".

La nueva realidad conformada por el Reglamento, no debe suponer una restricción en el desarrollo del mercado interior de la Unión, sino todo lo contrario, por ello el camino idóneo no es el establecimiento de restricciones en el tratamiento de datos de carácter personal, sino la búsqueda de otras fórmulas que refuercen las propias garantías jurídicas que deben presidir dichos tratamientos. Concretamente, se sugiere que se haga referencia a los conceptos de "control", y de "transparencia informativa", tan usados en protección de datos, y que son los que deben ser también aplicados a los tratamientos derivados de "Big Data".

Por tanto, con el fin de conciliar los legítimos derechos a la protección de datos de carácter personal con el desarrollo de la industria, es importante ponderar ambas situaciones en la necesidad de tutelar adecuadamente los respectivos intereses, atendiendo a conceptos como "transparencia" en los tratamientos, o "control" sobre los mismos antes citados.

Debe ponerse de manifiesto que la exigencia general del consentimiento podría en la práctica suponer un obstáculo desde el punto de vista de la ponderación con la que han de ser interpretadas las normas jurídicas, pues se impondrían unos límites desproporcionados a los desarrollos que se pretendan hacer de esta técnica.

Por ello, se debe 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. Por ello debe insistirse en que dicho conocimiento puede ser la piedra angular que supere la dialéctica sobre la necesidad del consentimiento e incluso del derecho de información del titular de los datos en la técnica Big Data, en el sentido jurídico previsto para los mismos en la legislación de protección de datos.

Por conclusión, cualquier uso previsto o futuro que se quiera llevar a cabo de los datos de carácter personal, necesariamente debe partir de la exigencia legal y moral del responsable del fichero y/o tratamiento de actuar transparentemente con el titular los datos.

II.3.3.- CALIDAD DE LOS DATOS Y CONSERVACIÓN.

Dado que el fenómeno del Big Data supone la gestión y tratamiento de ingentes cantidades de datos personales que, si bien es cierto, puede brindar grandes ventajas y beneficios a las organizaciones públicas y privadas y a la sociedad en general, también puede conllevar ciertos riesgos en materia de privacidad. Por ello, es importante que cada entidad ponga especial interés en cumplir, entre otros, con el principio de calidad de los datos personales utilizados en proyectos Big Data, así como en la conservación y almacenamiento de los mismos.

En primer lugar, la calidad de los datos en un entorno de Big Data debe estar siempre y, en todo momento, contemplada desde el inicio del proyecto ya que, en caso contrario, pueden surgir problemas de gran calado para la organización. Entre otros, podrían contemplarse los siguientes:

- Realizar tratamientos que son incompatibles con la finalidad que motivó su recogida.
- Gestionar datos desactualizados o erróneos que llevan a resultados incorrectos.
- Tomar decisiones o realizar tratamientos basados en datos inexactos que no responden a la realidad.
- Dificultar la gestión de derechos de acceso, rectificación, cancelación y oposición de los datos personales que trata la entidad.

En este sentido, el RGPD exige que todo tratamiento de datos personales cumpla con el principio de calidad:

- Que sean tratados de manera lícita, leal y transparente.
- Que sean datos pertinentes y adecuados a la finalidad que motiva la recogida.
- Que estén estrictamente limitados a los necesarios atendiendo a la finalidad para la que se recogen.
- Que no se utilicen para finalidades incompatibles o que nada tengan que ver con aquellas que motivaron su recogida.
- Que sean exactos y estén actualizados.

Por todo lo anterior, es recomendable que desde el momento que se sienten las bases para abordar un proyecto Big Data, la entidad tenga en cuenta lo siguiente:

- Que se realice un análisis previo al inicio del proyecto sobre la tipología de datos necesarios, las finalidades iniciales que se pretenden con la recogida de datos y aquellas que puedan surgir en un futuro cercano, y la caducidad de los mismos.

• Que las empresas sepan escoger los datos correctos e interpretarlos de manera adecuada (gestión de datos adecuados, pertinentes y no excesivos).

• Que las empresas organicen la información que obtienen de manera coherente, en función de los propósitos perseguidos para evitar recoger y almacenar más datos de los estrictamente necesarios para cumplir la finalidad para la que se recogieron.

• Que se establezcan protocolos de verificación periódica que permitan comprobar que el tratamiento de datos continúa siendo compatible y lícito con la finalidad inicial.

• Que la entidad abogue por cumplir con el principio de minimización de datos, es decir, limitar al mínimo la cantidad de información recabada que permita cumplir con el propósito legítimo que se pretende y que se almacene durante el tiempo mínimo indispensable. Así, es recomendable que la entidad establezca revisiones periódicas automáticas y técnicas de revisión y depuración de información.

En segundo lugar, la conservación de los datos (tanto desde el punto de vista de la forma en que se realiza su almacenamiento como respecto al periodo durante el que deben almacenarse) tiene también un papel trascendental a la hora de abordar un proyecto de Big Data.

El artículo 32 del RGPD, dedicado a la seguridad del tratamiento, no detalla las medidas de seguridad relacionadas con la conservación de la información, sino que deberán ser definidas por el responsable mediante un adecuado análisis de riesgos o evaluación de impacto.

Puede ser recomendable para la entidad establecer sistemas de control de acceso (por ejemplo, estableciendo protocolos de acceso específicos en función del tipo de usuario), realizar el cifrado de la información y, en la medida de lo posible, utilizar técnicas de monitorización del sistema que sean objeto de auditorías periódicas.

Respecto al periodo durante el que deberán almacenarse los datos, se limitará al tiempo imprescindible para cumplir con la finalidad que motivó la recogida de los datos y, en su caso, durante el tiempo que exija la normativa sectorial que resulte de aplicación. Por ello, puede ser recomendable que la entidad realice una división de datos en función del tiempo durante el que deben ser almacenados. Dichos criterios deberán ser definidos con el apoyo del departamento legal de la entidad.

II.3. 4.- DERECHOS DE LOS INTERESADOS.

La protección de datos de carácter personal no puede entenderse sin ofrecer a los interesados un conjunto de derechos que les garanticen el control del uso que se hace de sus datos personales por parte de los responsables y encargados de esos tratamientos.

En este sentido, en los artículos 15 y ss del RGPD se regulan los derechos de acceso, rectificación, supresión (derecho al olvido), limitación, portabilidad y oposición, así como el de decisiones individuales automatizadas.

Con todo ello, el sistema vigente de protección de derechos queda configurado de la siguiente manera:

• Derecho de acceso: el afectado tiene derecho a conocer si sus datos están siendo tratados, y entre otra información, tiene derecho a conocer las finalidades del tratamiento, las categorías de datos tratados, los destinatarios o categorías de destinatarios a los que se comunican o se pueden comunicar esos datos, plazos de conservación de los datos si es posible, origen de los datos cuando no hayan sido obtenidos de propio interesado o las transferencias internacionales previstas o realizadas. El derecho de acceso se entiende cumplido con la entrega de una copia de esa información por parte del responsable de tratamiento al interesado. Las copias adicionales pueden conllevar el cobro de un canon razonable por parte del responsable del tratamiento.

• Derecho de rectificación: derecho a que, sin dilación indebida, los datos sean rectificados cuando resulten en todo o en parte inexactos, así como que se completen los datos que no lo estén.

• Derecho de supresión (derecho al olvido): derecho a la supresión de los datos que le conciernen sin dilación indebida en, entre otros, los siguientes supuestos: los datos ya no son necesarios para la finalidad para la que fueron obtenidos, el interesado revoca el consentimiento habilitante del tratamiento y no existe otro fundamento jurídico que habilite el tratamiento, o el interesado ejercita su derecho de oposición para tratamientos de mercadotecnia o lo ejercita para oponerse a otros tratamientos como la elaboración de perfiles y no prevalezcan otros motivos legítimos para el tratamiento.

• Derecho de oposición: derecho del interesado a oponerse al tratamiento de sus datos inclusive los relacionados con la elaboración de perfiles cuando la finalidad sea la de mercadotecnia directa, así como el derecho del interesado a oponerse al tratamiento de sus datos cuando el tratamiento esté basado en un motivo o causa de interés legítimo o interés público o ejercicio de poderes públicos, incluida la elaboración de perfiles.

• Derecho a no verse sometido a un tratamiento basado únicamente en decisiones automatizadas individuales (artículo 22): derecho a no verse sometido a una decisión basada únicamente en un tratamiento automatizado, incluida la creación de perfiles, que produzca efectos jurídicos en él o le afecte significativamente de modo similar, salvo que el tratamiento se ampare en alguna de las excepciones previstas en el artículo citado.

Entre los nuevos derechos regulados en el RGPD, se encuentran el derecho de limitación del tratamiento (artículo 18), para supuestos tasados como, por ejemplo, en casos en los que es necesario realizar comprobaciones sobre la exactitud o inexactitud de los datos o cuando es necesario suspender el borrado de los datos porque así lo solicita el interesado para la formulación, ejercicio o defensa de reclamaciones; o el derecho a la portabilidad de datos (artículo 20) entendido como el derecho del afectado a obtener, en un formato estructurado, de uso común y lectura mecánica, la información que le concierna y haya facilitado a un responsable de tratamiento cuando esa información se trata por medios automatizados y sobre la base del consentimiento o para la ejecución de un contrato. En relación con este último derecho, recientemente el Grupo de Trabajo del Artículo 29 de la Directiva 95/46/CE ha adoptado directrices sobre la aplicación del derecho a la portabilidad. En este sentido, por ejemplo, el Grupo considera que el concepto de datos facilitados por el interesado incluye los datos proporcionados de manera activa por el interesado y los datos observados (datos de ubicación, búsqueda, ritmo cardiaco, etc) pero no incluye dentro de los datos sujetos al derecho a la portabilidad a los datos inferidos o deducidos que hayan sido creados por el responsable de tratamiento a partir de los datos proporcionados por el interesado (como pueden ser los resultados algorítmicos).

Trasladando el ejercicio de derechos a los tratamientos de Big Data y como ocurriera en los epígrafes relativos a información, consentimiento y calidad de los datos, la principal cuestión a tener en cuenta y resolver versa sobre los usos y tratamientos futuros y no previstos en el momento de la captación del dato.

Si en esos epígrafes la problemática se suscitaba en cómo poder mantener informado permanentemente al afectado sobre los nuevos usos y finalidades del tratamiento y el modo de obtención de los nuevos consentimientos o ampliación de los anteriores, en lo que respecta al ejercicio de derechos el foco de atención debe centrarse en el modo de facilitar información permanente a los afectados sobre el modo y procedimiento mediante el cual pueden ejercitar sus derechos, así como la manera en la que, una vez ejercitado cualquiera de los mismos, debe ser atendido por el responsable o encargado de tratamiento.

Así por plantear algunos supuestos, estos serían algunos aspectos a tomar en consideración antes de iniciar el tratamiento de Big Data:

• La información al afectado se facilita en el momento de captación del dato, mientras que los tratamientos de Big Data, si por algo se caracterizan, es por su continuidad en el tiempo. De ahí la necesidad de implantar un sistema que facilite información sobre el modo y el procedimiento para el ejercicio de derechos, y que la misma sea de fácil acceso y localización por parte de los afectados.

• En los tratamientos de Big Data es habitual combinar información procedente de diferentes fuentes, tanto endógenas como exógenas. Ello no puede ser obstáculo para el ejercicio de derechos y se debe estar en posición de poder facilitar tanto el origen de la información como las comunicaciones de datos realizadas a terceros y, en su caso, las transferencias internacionales de datos que se han efectuado, ya que esa trazabilidad permitirá dirigirse por parte de los afectados, si así lo estiman oportuno, a otros responsables y encargados de tratamiento.

• La disociación del dato personal no debe ser una excusa y una traba para cursar los derechos. Se debe estar en disposición de poder informar al afectado sobre el hecho de la anonimización, si se ha producido, y del riesgo de reidentificación existente en el caso del ejercicio de un derecho de acceso. Para el resto de los derechos se deberá estar a la posibilidad de reidentificación por parte del responsable o encargado de tratamiento a la hora de atender su ejercicio por parte del afectado. Sólo será aplicable la normativa de protección de datos si es posible la reidentificación del interesado, aunque ello no exonera del deber de responder a la solicitud advirtiendo del uso de datos anonimizados y, en su caso, del riesgo residual de reidentificación.

• Se deben utilizar soluciones tecnológicas actuales y adecuadas, que deben evolucionar conforme se vayan produciendo nuevos desarrollos tecnológicos.

10 Grupo de Trabajo de Artículo 29 de la Directiva 95/46/CE / Directrices sobre el derecho a la portabilidad de los datos / 13 de diciembre de 2016.

II.3.5.- DECISIONES INDIVIDUALES AUTOMATIZADAS.

El RGPD contempla tres aspectos sobre la regulación de las decisiones automatizadas que tendrán una gran trascendencia en cualquier proyecto de Big Data.

En primer lugar, los principios de protección de datos, según el considerando 26 del RGPD, "no deben aplicarse a la información anónima, (…), ni a los datos convertidos en anónimos de forma que el interesado a quien se refieren no sea, o ya no resulte, identificable. En consecuencia, el presente Reglamento no afecta al tratamiento de dicha información anónima, ni siquiera con fines estadísticos y de investigación".

En segundo lugar, el RGPD da mucha importancia a los principios relativos al tratamiento leal y transparente de los datos, debiendo facilitar al interesado cuanta información complementaria sea necesaria a estos fines, así como de la existencia de las elaboraciones de perfiles y de sus consecuencias. Dicha información, tal y como señala el considerando 60 del RGPD, "… puede transmitirse en combinación con unos iconos normalizados que ofrezcan, de forma fácilmente visible, inteligible y claramente legible, una adecuada visión del conjunto del tratamiento de datos previsto".

El artículo 22 del RGPD regula una variante del derecho de oposición respecto de las decisiones individuales automatizadas, incluida la elaboración de perfiles, que va a tener una gran importancia en los tratamientos de Big Data. Según el artículo 4.4 del RGPD, elaboración de perfiles es toda forma de tratamiento automatizado de datos personales consistente en utilizar datos personales para evaluar determinados aspectos personales de una persona física, en particular para analizar o predecir aspectos relativos al rendimiento profesional, situación económica, salud, preferencias o intereses personales, fiabilidad o comportamiento, ubicación o movimientos de dicha persona física. Por tanto es importante destacar que 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.

Ello supondrá que el interesado podrá oponerse a dicho tratamiento, pero siempre y cuando: a) la decisión "automatizada" o basada en profiling produzca efectos jurídicos que conciernan al interesado (v.gr. un banco que decide no conceder un préstamo o una hipoteca basándose en determinados perfiles de riesgo) o b) la decisión "automatizada" o basada en profiling le afecte significativamente.

En este aspecto es importante tener en cuenta que la premisa básica es que exista una decisión exclusivamente "automatizada" sin que medie intervención humana alguna, debiendo informarse al interesado acerca de la existencia de un mecanismo de decisión automatizado que comprenda la elaboración de perfiles, la lógica aplicada, importancia y consecuencias para el interesado.

Cabe añadir que el citado artículo recoge excepciones a lo anterior y señala que lo anteriormente citado no se aplicará si la decisión es necesaria para la celebración o la ejecución de un contrato entre el interesado y un responsable del tratamiento o se base en el consentimiento explícito del interesado.

En tercer lugar, el último apartado del artículo 22 del RGPD establece una prohibición general de adoptar decisiones individualizadas automatizadas basadas en datos personales sensibles (origen étnico o racial, opiniones políticas, convicciones religiosas o filosóficas, afiliación sindical, datos genéticos, datos biométricos o datos relativos a la salud, vida y orientación sexuales) salvo consentimiento explícito del interesado o por motivos de interés público, siempre que se hayan tomado medidas adecuadas para salvaguardar los derechos y libertades, y los intereses legítimos del interesado.

III.PRINCIPIOS Y ASPECTOS PROCEDIMIENTALES.

III.1.- PRIVACIDAD DESDE EL DISEÑO.

Una de las buenas prácticas a tener en cuenta en todo proyecto relacionado con Big Data es precisamente considerar la privacidad desde el diseño, con el objetivo de asegurar que las garantías de protección de los datos se incorporan ya desde la fase de planificación de los procedimientos y sistemas de información.

En la práctica, supone tener en consideración la privacidad y el cumplimiento de las normativas de protección de datos desde la fase inicial del proyecto (de la misma manera que se tienen en consideración el resto de requisitos funcionales y no funcionales) con el objetivo de que el proyecto se diseñe e incluso ajuste y desarrolle teniendo en consideración dichos requerimientos, de tal manera que la privacidad se integre en las nuevas tecnologías y prácticas empresariales directamente, desde el principio, como un componente esencial de la protección de la privacidad.

Por otra parte, no hay que perder de vista que, si se tienen en consideración estos aspectos desde el inicio, se evitará tener que redefinir los sistemas y procesos continuamente y, por lo tanto, incurrir en costes futuros asociados a la implantación de estos requerimientos.

En este sentido, uno de los enfoques más ampliamente reconocidos relacionado con la privacidad proactiva es Privacy by Design (PbD), concepto desarrollado en la década de los noventa y es el que se está adoptando globalmente por un creciente número de organizaciones y jurisdicciones.

El concepto protección de datos desde el diseño y por defecto, recogido en el artículo 25 del RGPD, consiste en incorporar, desde las primeras fases de todo proyecto, medidas técnicas y organizativas apropiadas, teniendo en cuenta factores como el estado de la técnica, el coste de la aplicación o los riesgos del tratamiento para los derechos y libertades de los afectados, para cumplir los requisitos del Reglamento y proteger los derechos de los interesados.

Para poder llevarlo a la práctica, se establecen los siguientes 7 principios fundamentales:

- Proactivo no reactivo; preventivo no correctivo: la privacidad desde el diseño suele caracterizarse por tomar medidas proactivas en lugar de reactivas. Se anticipa y previene la pérdida de privacidad de la información antes de que suceda.

- La privacidad como Configuración por Defecto o Privacidad por Defecto: ofrecer el máximo grado de privacidad para asegurar que los datos personales están protegidos automáticamente en cualquier sistema informático o dentro de las buenas prácticas. Sin necesidad de actuación por parte del cliente o proveedor, la protección de su información y su privacidad se mantiene intacta, ya que está integrado en el sistema por defecto.

- La privacidad embebida en el diseño: la protección de la información debe estar embebida en la infraestructura TI y en los procesos de la empresa. No debe ser considerado como un añadido sino como un componente esencial del núcleo como parte integral del sistema, sin disminuir la funcionalidad.

- Funcionalidad completa- Suma-Positiva, no de Suma-Cero: con este principio se pretende dar cabida a todos los intereses y objetivos legítimos de una forma de suma positiva “win-win”, no a través de un enfoque anticuado de suma cero, donde se hacen innecesarias las compensaciones. Se trata de garantizar que se cubren todas las funcionalidades y necesidades de los distintos implicados, pero sin afectar a la privacidad. Privacidad desde el diseño evita la pretensión de falsas dicotomías, como la privacidad frente a la seguridad. No tiene sentido pensar en la privacidad sin la seguridad ni la seguridad sin la privacidad.

- Seguridad punto-a-punto- Protección completa del ciclo de vida de los datos: desde el momento de su recolección, la protección se extiende a través de todo el ciclo

de vida de los datos involucrados. De esta manera, todos los datos se conservan y destruyen de forma segura, asegurando la gestión del ciclo de vida seguro de la información, punto a punto.

- Visibilidad y transparencia- Mantenerlo abierto: garantizar a todos los interesados que, sean cuales sean las prácticas de negocio o la tecnología utilizadas, funcionarán de acuerdo con los compromisos y los objetivos establecidos, y que estarán sujetos a una verificación independiente. De esta forma los componentes y operaciones permanecen visibles y transparentes, a los usuarios y proveedores por igual. Recuerda: ¡Confía, pero verifica!

- El respeto a la privacidad del usuario- Manténgala centrada en el usuario: por encima de todo, la privacidad desde el diseño requiere que los desarrolladores y operadores del sistema mantengan por encima de todo el interés de las personas, ofreciendo unas medidas de protección fuertes en sus valores predeterminados de privacidad, con avisos apropiados, y fortalecer las opciones para que sean fáciles de usar.

Si en todo proyecto de Big Data se tuviesen en consideración estos principios, y muy en particular los asociados a la Limitación en la Recogida y la Minimización de Datos, se reduciría considerablemente el riesgo para la privacidad.

III.2.- "ACCOUNTABILITY".

El principio de accountability está íntimamente relacionado con la responsabilidad social corporativa o institucional vinculado al desarrollo de las nuevas tecnologías, especialmente en lo atinente a los tratamientos de datos de carácter personal que se puedan llevar a cabo.

Por ello, puede afirmarse que la accountability constituye un principio consistente en el reconocimiento, asunción de responsabilidad y actitud transparente sobre los impactos de las políticas, decisiones, acciones, productos y desempeño asociados a una organización.

La accountability constituye una filosofía que implica la procedencia de dar cumplimiento al régimen jurídico y las obligaciones derivadas de la protección de datos de carácter personal, con independencia de que exista una norma concreta de carácter imperativo que así lo exija. Por ello, esto obliga a las organizaciones a implicar a los grupos de interés para identificar, comprender y responder a los temas y preocupaciones existentes en este ámbito, a los efectos de poder garantizar adecuadamente la sostenibilidad jurídica y social de los tratamientos de datos, informando, explicando y dando repuesta al efecto al regulador, a los ciudadanos como titulares de los datos, y a la sociedad en general acerca de las decisiones, las acciones y el desempeño.

En materia de protección de datos este principio alude, tal como señala la Agencia Española de Protección de Datos, a la responsabilidad de las compañías en la implantación de medidas, en el seno de sus organizaciones, de garantía y cumplimiento de los principios y obligaciones en materia de protección de datos, así como el establecimiento de mecanismos internos y externos para evaluar su fiabilidad y demostrar su efectividad cuando se solicite por las autoridades de control.

Este principio tiene una gran relevancia tanto en entornos públicos como privados, particularmente en el contexto actual, marcado por el empleo intensivo de nuevas tecnologías y, fundamentalmente, de los servicios de Internet.

Consecuencia de todo lo anterior, el Grupo de Trabajo del Artículo 29 considera que las medidas comunes de responsabilidad, entre las que se puede materialmente concretar este principio de la accountability, pueden ser, entre otras, las que se citan a continuación:

- El establecimiento de procedimientos internos previos a la creación de nuevas operaciones de tratamiento de datos personales (revisión interna, evaluación, etc.).

- El establecimiento de políticas escritas y vinculantes de protección de datos que se tengan en cuenta y se valoren en nuevas operaciones de tratamiento de datos (p.ej., cumplimiento de los criterios de calidad de datos, notificación, principios de seguridad, acceso, etc.) que deben ponerse a disposición de las personas interesadas.

- La cartografía de procedimientos que garanticen la identificación correcta de todas las operaciones de tratamiento de datos y el mantenimiento de un inventario de las mismas.

- El nombramiento de un delegado de protección de datos.

- La oferta adecuada de formación en protección de datos a los miembros del personal; esto debe incluir a los responsables de los procesos de datos personales (como los directores de recursos humanos), pero también a los administradores de tecnologías de la información, desarrolladores en general, y directores de unidades comerciales. Deben asignarse recursos suficientes para la gestión de la privacidad, el establecimiento de procedimientos de gestión del acceso y de las demandas de corrección y eliminación de datos con transparencia para las personas interesadas.

• El establecimiento de un mecanismo interno de resolución de quejas de los interesados. En este ámbito puede jugar un papel destacado el Delegado de Protección de Datos.

• El establecimiento de procedimientos internos de gestión y notificación eficaces de fallos de seguridad (violaciones de seguridad).

• La realización de evaluaciones de impacto sobre la privacidad en circunstancias específicas.

• La aplicación y supervisión de procedimientos de verificación que garanticen que las medidas no sean sólo nominales, sino que se apliquen y funcionen en la práctica (auditorías internas o externas).

Por otra parte, el Reglamento promueve la elaboración de códigos de conducta como mecanismos o instrumentos de autorregulación para garantizar su cumplimiento, previendo expresamente que establezcan procedimientos extrajudiciales de resolución de conflictos que permitan resolver las controversias entre los responsables del tratamiento y los interesados.

En definitiva, y a los efectos que nos ocupan, cualquier planteamiento de Big Data debe ajustarse a los principios de la accountability.

El GT29 en sus opiniones está convencido de que el aumento no sólo de los riesgos sino del valor de los datos personales en sí abunda en la necesidad de reforzar el papel y la responsabilidad de los responsables del tratamiento de datos.

El artículo 5.2 del RGPD recoge el principio de accountability (responsabilidad proactiva), es decir, que el responsable del tratamiento debe estar en disposición de demostrar que cumple con lo regulado en el Reglamento.

III.3.- EVALUACIÓN DE IMPACTO (EIPD).

La Evaluación de Impacto de la Protección de Datos, en adelante EIPD, es un proceso que ha de permitir a las empresas y administraciones determinar si las iniciativas que involucran el uso de información privada representan riesgos para el derecho a la protección de datos y, como valor añadido, les permite medir, cuantificar dichos riesgos y valorar el impacto que tienen sobre los derechos y libertades de las personas cuyos datos personales tratan.

Por su parte, el Reglamento Europeo de Protección de Datos formula un modelo de cumplimiento basado en la gestión enfocada al riesgo de manera que la EIPD se constituye como una herramienta clave para garantizar la privacidad de productos y servicios, puesto que sirve para poder justificar y evaluar correctamente las decisiones que se adopten y que impliquen la realización de cualquier clase de tratamiento.

Big Data se caracteriza por incorporar en el *Business Intelligence* o inteligencia empresarial las fuentes de Internet (redes sociales, blogs, foros, medios de comunicación…), la actualización permanente de las mismas y el carácter continuo e inmediato de los análisis, hechos que exponen y elevan los riesgos potenciales para la privacidad.

En este sentido, las organizaciones deben ser especialmente cautas con los riesgos asociados a sus procesos de identificación, análisis y recolección de información.

Al adoptar nuevas soluciones tecnológicas como Big Data, todos los riesgos deben ser identificados y gestionados. Eso incluye desarrollar un sistema de administración y gestión de los mismos acorde con la estructura organizativa y los procesos relacionados con tratamientos de datos personales, que garantice la continuidad de los procesos, así como hacer frente, entre otros, a los riesgos legales y regulatorios.

En el RGPD se exige la realización de evaluaciones de impacto como herramientas indispensables para evaluar “el origen, naturaleza, particularidades y gravedad del riesgo, en los casos en que las operaciones de tratamiento puedan dar lugar a un alto riesgo para los derechos y libertades de las personas”. Concretamente la propuesta del RGPD, establece la necesidad de realizar la EIPD siempre que se den las siguientes circunstancias:

• Cuando las operaciones de tratamiento impliquen llevar a cabo una evaluación sistemática y amplia de aspectos personales relativos a personas físicas, que incluye la elaboración de perfiles, y especialmente si sobre el resultado del tratamiento se basan decisiones que produzcan efectos jurídicos sobre el individuo, o pueden afectar de manera significativa a los individuos.

• El tratamiento a gran escala de datos sensibles, es decir, los referidos en el artículo 9 del RGPD: los que revelen el origen étnico o racial, opiniones políticas, convicciones religiosas o filosóficas, afiliación sindical, tratamiento de datos genéticos, datos biométricos dirigidos a identificar de manera unívoca a una persona y datos relativos a la salud, la vida sexual o la orientación sexual de una persona.

• Los datos obtenidos del control de áreas de acceso público a gran escala, mediante monitorización por sistemas de video vigilancia.

Bajo estas circunstancias, las organizaciones deberán incluir en sus proyectos Big Data relacionados con la recopilación, uso y divulgación de datos personales una EIPD que venga a demostrar que dicho tratamiento pretende lograr un objetivo específico y legítimo, y que el riesgo en la privacidad una vez adoptadas las medidas de seguridad necesarias, sea residual.

El artículo 35 del RGPD establece el contenido mínimo que debe incluir una evaluación de impacto. Un guion adecuado para la confección de la EIPD, podría ser el sugerido por la AEPD en la guía publicada en su página web, que contiene los puntos siguientes:

• Descripción sistemática del/los tratamiento/s considerados. Se trata de conocer la tipología de datos a tratar, los soportes para el tratamiento, duración prevista, las tecnologías de información involucradas y sus aspectos funcionales, el flujo de datos, los destinatarios (...), en una descripción detallada del ciclo de vida en el tratamiento de los datos.

Dado que los proyectos Big Data afectan a un volumen importante de personas se han de llevar a cabo consultas a las partes interesadas, e incluir dichas contribuciones en el documento de evaluación.

Cuanto más exhaustiva es esta primera descripción más puede ayudar a las organizaciones a considerar de inicio métodos menos invasivos, ya sea recogiendo datos anónimos o bien utilizando tecnologías respetuosas con la confidencialidad.

• Identificación y valoración de los riesgos. En función del contexto de tratamiento de datos, hay que identificar las fuentes de riesgos, los posibles escenarios de sucesos no deseados y amenazas que los pueden hacer posibles, las vulnerabilidades que pueden facilitar la materialización de dichas amenazas, así como los impactos o consecuencias posibles sobre la vida privada de las personas afectadas.

Los riesgos a valorar podrían clasificarse, en función de si afectan a las personas cuyos datos son tratados o si afectan a la organización que trata dichos riesgos. En el primer caso nos encontramos con impactos tales como la vulneración de los derechos de las personas, la pérdida de la información tratada o el daño causado por el uso incorrecto de los datos.

Los impactos para las organizaciones podrían resumirse en incumplimientos legales, pérdida de reputación, así como la posibilidad de acciones sancionadores o de responsabilidad.

La estimación de los riesgos, finalmente se realiza en términos de severidad de los impactos y probabilidad de ocurrencia de los riesgos, considerando en esta valoración la ponderación en el riesgo por las medidas de seguridad existentes.

• Gestión de los Riesgos evaluados y selección de las medidas de seguridad que permitan reducir los riesgos o su impacto final. La EIPD debe recoger la descripción de las medidas tanto organizativas como de seguridad lógica y física, y su influencia en la disminución de la probabilidad y las consecuencias.

• Análisis del cumplimiento normativo. Una vez valorados los riesgos, las medidas previstas y el impacto del cumplimiento normativo, los tratamientos deben asegurar el cumplimiento de los requisitos legales establecidos en la legislación vigente.

• Informe final y conclusión. De resultas de estudiar las etapas precedentes, la EIPD debe contener, a modo de conclusión, las recomendaciones con las medidas que deben adoptarse bien sean de eliminación, mitigación, transferencia o aceptación de los riesgos para la privacidad.

• Implantación de las recomendaciones. Para asegurar la efectiva implantación de las medidas identificadas por la evaluación de impacto es necesario asignar los recursos necesarios y verificar el seguimiento de todas las fases del proceso y la correcta implantación de las medidas en relación con los objetivos establecidos para tratar los riesgos de privacidad.

• Revisión y realimentación. La organización debe establecer un plan de supervisión y revisión que permita auditar los resultados de la evaluación de impacto y las medidas adoptadas en aplicación de los mismos, incluyendo estas revisiones como un elemento más en su gestión empresarial.

Estas revisiones deberán realizarse periódicamente, siempre que aparezcan nuevos riesgos o cuando las condiciones de tratamiento se modifiquen de manera substancial, ya sea por la aparición de nuevas tecnologías, nuevos afectados, nuevos datos, etc.

Asimismo deben tenerse en cuenta las directrices elaboradas en abril de 2017 por el GT29 sobre evaluaciones de impacto y para determinar si un tratamiento es susceptible de producir un alto riesgo en relación con los objetivos del RGPD.

Las directrices explican el sentido de los artículos 35 y 36 del RGPD y aportan, entre otros, los siguientes elementos:

- Una lista común para la UE de operaciones de tratamiento de datos para los cuales la PIA es obligatoria con arreglo al art. 35.4.
- Una lista de tratamientos para los cuales las PIAs no resultan necesarias con arreglo al art. 35.5.
- Criterios comunes sobre la metodología para las evaluaciones de impacto con arreglo al mencionado art. 35.5.

Por último la guía da respuesta, en cuatro apartados, a asuntos relevantes en relación con las evaluaciones:

1.- Cuál es el objeto de la PIA: operaciones únicas y conjuntos de operaciones similares de tratamientos de datos.
2.- Cuándo se debe someter de forma obligatoria un tratamiento de datos a evaluación PIA.
3.- Cómo llevar a cabo una PIA.
4.- Cuándo se deberá consultar a la autoridad de supervisión (en el caso de que el riesgo residual para la protección de datos sea elevado).

III.4.- REUTILIZACIÓN DE DATOS DISOCIADOS.

Para garantizar la irreversibilidad de los procesos de anonimización de los datos personales utilizados en las iniciativas de Big Data y, de este modo, la no aplicación de la normativa de protección de datos, se habrán de considerar tanto las fuentes de información disponibles en los diferentes medios, especialmente en internet, como la tecnología disponible, no solo por parte del responsable del tratamiento sino por cualquier otra persona.

La AEPD, consciente de las dificultades para encontrar el ajuste perfecto entre anonimización e irreversibilidad, ha elaborado unas pautas útiles para implantar estas técnicas que se habrán de tener en cuenta ante procedimientos de anonimización de datos personales en iniciativas de Big Data.[12]

En este mismo sentido, el GT29 analiza en el Dictamen 5/2014[13], los límites que deben considerarse a la hora de aplicar procesos y técnicas de anonimización de datos personales.

Según lo indicado por el GT29, para que exista una verdadera anonimización de datos personales, ésta debe ser irreversible, es decir, que razonablemente no permita la identificación del titular de los datos personales, si bien es preciso valorar los riesgos derivados de las técnicas implementadas para dicha anonimización, ya que dado el estado actual de la tecnología, pueden suscitarse situaciones que, aunque a priori no parecen permitir la reversibilidad, mediante la aplicación de ciertas tecnologías, sí se podría llegar a identificar al interesado.

En virtud de lo anterior, conforme a lo indicado por el GT29, es necesario realizar una evaluación de las técnicas y procedimientos de anonimización utilizados, a fin de acreditar que la disociación realizada evita que: (i) se pueda identificar a una persona física dentro de un conjunto de datos; (ii) se pueda relacionar o enlazar la información de una persona física a partir de la vinculación de dos registros dentro de un conjunto de datos (o entre dos conjuntos de datos independientes); o (iii) se pueda inferir cualquier información sobre la persona física en un conjunto de datos.

En este sentido, el GT29 advierte que los datos incluso anónimos, como las estadísticas, se pueden usar para enriquecer los perfiles existentes de individuos, creando así nuevos supuestos de protección de datos. Por lo tanto, la anonimización no debe ser considerada como un ejercicio aislado, y los riesgos existentes se deben reevaluar regularmente por los responsables del tratamiento.

En este mismo sentido, la AEPD establece en sus Orientaciones sobre procedimientos de anonimización de datos personales que en un proceso de anonimización es fundamental poder valorar los riesgos de reidentificación posterior y cómo se van a garantizar los derechos de las personas en tal caso.

La AEPD resalta, en cuanto a la selección de las técnicas de anonimización, la utilidad de los algoritmos de cifrado para este tipo de procesos, resaltando los algoritmos de "hash" como fórmula para garantizar la confidencialidad del dato por tratarse de una operación en un solo sentido, es decir, partiendo de un dato podemos generar siempre la misma huella digital, pero partiendo de una huella digital nunca podremos obtener el dato original. No obstante, aclara que un mecanismo de hash no

11 Orientaciones sobre protección de datos en la reutilización de la información del sector público.
12 Orientaciones y garantías en los procedimientos de anonimización de datos personales.
13 Dictamen 5/2014 sobre las técnicas de anonimización de datos personales.

garantiza por sí solo la irreversibilidad del dato, es preciso combinarlo con otras medidas tales como la aplicación de algoritmos de cifrado, la utilización de sellos de tiempo, o la aplicación de capas de anonimización, atendiendo a la criticidad de las variables de identificación, a la organización interna de quien ejecuta el tratamiento y a su política de anonimización.

En combinación con las técnicas de anonimización mencionadas, pueden utilizarse otras como las de perturbación de datos o los métodos de reducción de datos.

Basándose en lo anterior, para determinar qué técnica de anonimización debe ser aplicada, se deberá partir siempre de cuál es la finalidad que se busca en relación al proceso de anonimización de los datos, realizando inicialmente el correspondiente análisis de riesgos del proceso de anonimización para posteriormente gestionar los riesgos resultantes con medidas técnicas, organizativas o de cualquier otra índole.

Conviene en este punto recordar que las técnicas de anonimización no garantizan en términos absolutos la imposibilidad de reidentificación, por lo que existirá siempre un índice de probabilidad de reidentificación que se deberá intentar atenuar mediante la correspondiente gestión de riesgos. Con este objetivo, aunque no sea resulte obligatorio, aplicar la metodología de la Evaluación de Impacto en la Protección de Datos siempre es muy recomendable y, en todo caso, será preciso llevar a cabo un proceso de evaluación de riesgos centrado en analizar las posibilidades de reidentificación de los interesados.

Igualmente, indica la autoridad de control que será necesario prever una hipotética pérdida de información por negligencia del personal implicado, por falta de una política de anonimización adecuada o por una revelación de secreto intencionada que diera lugar a la pérdida de las variables de identificación o claves de identificación de las personas.

Los procesos de anonimización se deben de enfocar desde el concepto de Protección de Datos desde el Diseño, lo que significa que los requisitos de privacidad deben ser tenidos en cuenta desde las etapas iniciales del diseño de las iniciativas de Big Data, para el proceso de anonimización y durante todo el ciclo de vida de dichas iniciativas.

Asimismo, debe tenerse en cuenta que, para determinar las técnicas y procedimientos con mayor idoneidad para realizar la anonimización, se deberá analizar cada caso en concreto, teniendo en consideración que, incluso un conjunto de datos anónimos, todavía podría presentar riesgos residuales a los titulares de los mismos.

Igualmente, será aconsejable definir una política de anonimización que se encuentre documentada y actualizada, de manera que refleje de forma justificada las actuaciones llevadas a cabo para proteger la privacidad de los interesados y se encuentre accesible al personal implicado en el tratamiento de datos anonimizados; así como un protocolo de actuación del proceso de anonimización que contemple, al menos, los siguientes parámetros o elementos:

- Identificación de activos implicados en el proceso de anonimización.

- Equipo de trabajo asignado y segregación de funciones, atendiendo a perfiles o roles en relación con el proceso de anonimización, en línea con el principio de independencia profesional.

- Realización de una EIPD teniendo en cuenta lo indicado en la Guía para la evaluación de impacto en la protección de datos personales publicada por la AEPD (identificación de riesgos, valoración de los riesgos existentes, salvaguardas encaminadas a evitar que los riesgos se materialicen, cuantificación del impacto de la posible materialización de los riesgos, informe de los riesgos resultantes, determinación del umbral de riesgo aceptable, gestión de los riesgos asumibles, informe final de los riesgos existentes y medidas a implantar para minimizar su impacto).

- Revisión de riesgos en caso de cambios en los procesos de anonimización y reevaluaciones periódicas del riesgo residual existente con el objetivo de introducir parámetros de mejora de la calidad de los procesos de anonimización.

• Formación e información al personal implicado en los procesos de anonimización con respecto al cumplimiento de la normativa de protección de datos personales, especialmente en relación con las medidas de seguridad de índole técnica y organizativa, la existencia y aplicación de una política de anonimización, medidas de control del personal con acceso a la información anonimizada, obligaciones y deberes en caso de ruptura de la cadena de anonimización y las actuaciones que debe realizar para paliar el impacto resultante de la materialización de alguno de los riesgos de reidentificación.

• Eliminación o reducción de variables que permitan la identificación de las personas cuyos datos se traten en las iniciativas de Big Data.

• Auditoría del proceso de anonimización y del uso posterior de los datos.

La política de anonimización, el protocolo de actuación y las medidas tecnológicas adoptadas respecto de los procedimientos de anonimización habrán de reforzarse con las garantías jurídicas necesarias para preservar los derechos de los interesados, tales como (i) acuerdos de confidencialidad y cláusulas contractuales que garanticen la privacidad de la información incluso cuando haya brechas de reidentificación; (ii) compromisos de mantenimiento de la anonimización de la información suscritos con los posibles destinatarios de la misma así como de no realizar ninguna acción para re-identificarla; o (iii) auditorías de uso de la información anonimizada.

Estas garantías serán tenidas en cuenta como parte de las salvaguardas adoptadas para la minimización de los daños ante una eventual reidentificación de los interesados.

En último lugar, se recomienda la realización de un proyecto piloto con una pequeña muestra de datos de prueba (no reales) del que puedan extraerse, de forma objetiva, conclusiones con respecto a la vialidad de las técnicas de anonimización propuestas y del procedimiento de anonimización.

III.5.- RELACIONES CON LA AUTORIDAD DE CONTROL.

Las autoridades nacionales de control juegan un papel primordial en la protección y la garantía del derecho fundamental a la protección de datos de carácter personal, debiendo éstas velar por el mismo, tal y como reconoce el propio TJUE[14], en aplicación de las facultades que aquéllas ostentan en virtud de la Carta de los Derechos Fundamentales de la Unión Europea y del nuevo paquete regulatorio europeo en este ámbito, en particular, contenidas en el RGPD. En este sentido, el RGPD reconoce de forma expresa que el establecimiento en los Estados Miembros de autoridades de control capacitadas para desempeñar sus funciones y ejercer sus competencias con plena independencia constituye un elemento esencial de la protección de las personas físicas con respecto al tratamiento de sus datos de carácter personal.

Por este motivo, cuando un responsable de tratamiento planea realizar un proyecto Big Data, éste debe contemplar en caso necesario la relación e interlocución con la correspondiente autoridad de control a estos efectos, tanto en el momento de diseño inicial del citado proyecto, cuanto durante su realización y desarrollo efectivo, considerando de forma especial los siguientes aspectos en función del momento al que se refiera, a saber:

A) Aspectos de interés a considerar en el momento del diseño:

A. 1. Consulta previa a la autoridad de control en base a los resultados de la Evaluación de Impacto en la Protección de Datos (EIPD) que se realice:

Según el art. 35 del RGPD, será la autoridad de control competente la que establezca la lista de operaciones o proyectos que requieran de una EIPD.

El resultado del análisis de riesgos podría entrañar un alto riesgo para los derechos o libertades de las personas físicas, lo que obligaría a realizar una evaluación de impacto en la protección de datos personales (EIPD). Así se recoge en la Guía publicada por la AEPD, donde ya se apunta como recomendable hacer una EIPD en los casos en que se traten grandes volúmenes de datos personales a través de tecnologías de datos masivos (Big Data).

Si el tratamiento de datos, en base a los resultados derivados de la EIPD, entraña un alto riesgo en caso de que no se adopten medidas para mitigarlo, el responsable debe consultar a la autoridad de control antes de proceder al mismo en los términos del artículo 36 del RGPD. Cuando la autoridad de control considere que el tratamiento previsto podría infringir la normativa aplicable deberá, asimismo, en un plazo de ocho semanas desde la solicitud de la consulta, asesorar por escrito al responsable, pudiendo prorrogarse dicho plazo por seis semanas más, en función de la complejidad del tratamiento previsto.

14 Entre otros pronunciamientos jurisprudenciales acerca de la importancia de las funciones y misión actual de las autoridades de control en el contacto mundial se destaca, en particular, el que sigue: http://curia.europa.eu/jcms/upload/docs/application/pdf/2015-10/cp150117es.pdf

A. 2. Adopción de medidas técnicas y organizativas apropiadas en proyectos Big Data:

Con carácter general, según prevé el artículo 24.1 del RGPD, la empresa o el responsable de tratamiento aplicará medidas técnicas y organizativas apropiadas a fin de garantizar y poder demostrar que el tratamiento es conforme con la normativa aplicable, y lo hará teniendo en cuenta la naturaleza, el ámbito, el contexto y los fines del tratamiento, así como los riesgos de diversa probabilidad y gravedad para los derechos y libertades de las personas físicas, debiendo considerar de forma especial los principios de privacidad por diseño y por defecto en este ámbito.

Además de planificar y adoptar las anteriores medidas, el responsable, en el marco de la consulta previa a la que alude el apartado anterior, debería informar de forma específica a la autoridad de control sobre las concretas medidas y garantías que estime adoptar en proyectos Big Data, debiendo la autoridad de control informar y asesorar al responsable en caso de que considere que éste no ha identificado o mitigado suficientemente el riesgo con las medidas que haya proyectado. Todo ello en coherencia con lo previsto en el artículo 57.1 del RGPD.

B) Aspectos de interés a considerar en el desarrollo:

B.1. Registro de actividades de tratamiento:

En cumplimiento del artículo 30 del RGPD, resulta obligatorio que el responsable, en el caso de desarrollar un proyecto Big Data, cuente con un Registro de las actividades de tratamiento asociadas al mismo, el cual, quedará a disposición de la autoridad de control competente.

Este registro será obligatorio para entidades con más de 250 trabajadores, u empleando a menos personas, en el caso de que el tratamiento proyectado pueda entrañar un riesgo para los derechos y libertades de los interesados (y ya se ha visto que los proyectos Big Data pueden comportar este riesgo), no sea ocasional, incluya categorías especiales de datos personales indicadas en el artículo 9, apartado 1 del RGPD, o datos personales relativos a condenas e infracciones penales a que se refiere el artículo 10 del mismo Reglamento.

B.2. Notificación por el responsable de tratamiento de una posible violación de la seguridad de los datos personales:

En caso de que se produzca una violación de la seguridad de los datos vinculados a un proyecto Big Data, deberá comunicarse la misma a la autoridad de control que corresponda según lo dispuesto en el art.33 del RGPD.

B.3. Coordinación del Delegado de Protección de Datos que se nombre con la autoridad de control:

A tenor de lo previsto en el artículo 37 del RGPD, es importante que el responsable de tratamiento nombre un Delegado de Protección de Datos (DPD) cuando vaya a realizar proyectos Big Data (al poder ser calificados como tratamientos a gran escala) y

que ejercerá las funciones dispuestas en el artículo 39 del RGPD y, con ello, el DPD nombrado será quien actúe como punto de contacto con la autoridad de control debiendo cooperar con ella. Del mismo modo, será quien brinde el asesoramiento que se le solicite por parte del responsable acerca de la evaluación de impacto relativa a la protección de datos que se realice sobre el proyecto Big Data, debiendo también supervisar su aplicación de conformidad con el artículo 35 del RGPD.

B.4. Deber general de cooperación con la autoridad de control en el desempeño de sus funciones (art.31 RGPD):

Al margen de lo anterior, existe para todo tipo de tratamientos y de proyectos sobre datos personales (sean o no de Big Data) un deber del responsable de tratamiento de colaborar con la autoridad de control en el correcto ejercicio de sus competencias, operando este deber como una garantía general en favor de la protección de la privacidad de las personas físicas en este tipo de proyectos.

IV. MEDIDAS TECNOLÓGICAS PARA LA MEJORA DE LA PRIVACIDAD, SEGURIDAD Y CONFIANZA.

La confianza depositada por el interesado en las empresas que dan soporte a este mundo digital es crucial para el aprovechamiento mutuo de los beneficios que reporta el Big Data.

Un estudio[15] reciente sobre la actitud de los europeos frente al potencial que el Big Data tiene para mejorar su vida o la de la sociedad revela que el 32% de los encuestados mantiene que aporta más ventajas que desventajas, mientras que un 51% piensa lo contrario. El estudio concluye que las reticencias de los encuestados desaparecen, en gran medida, cuando entienden de forma clara los beneficios que se derivan del uso de Big Data para ellos y para el conjunto de la sociedad.

Se pone así de manifiesto la necesidad que las organizaciones, públicas y privadas, tienen de combatir este escepticismo del interesado mediante una explicación clara de cómo analizan sus datos para que le transmita la confianza necesaria en la ERA Digital.

Por lo tanto, está claro que la confianza de los ciudadanos es clave para que pueda existir un despliegue de las potencialidades de las herramientas analíticas y, para que se produzca esa confianza, es imprescindible que las personas estén convencidas de que se ha tomado en serio su derecho a la privacidad y a la protección de datos; que se han evaluado los riesgos para su privacidad desde el inicio del proyecto y que se han puesto en marcha las medidas necesarias para eliminarlos o mitigarlos; y que en todo momento los tratamientos de datos en el área de Big Data se ajustan a la normativa de protección de datos en vigor.

IV.1.- ESTRATEGIAS DE PRIVACIDAD.

El concepto de privacidad y protección de datos desde el diseño es fundamental para hacer frente a los riesgos para la privacidad en las diferentes etapas de la cadena de valor de los proyectos de Big Data.

En este sentido, según un estudio sobre la privacidad por diseño y análisis de datos[16], existen diferentes estrategias de privacidad que pueden adoptarse desde el diseño que se exponen a continuación:

- Minimizar: La cantidad de datos personales debe limitarse lo máximo posible (minimización de datos).

- Agregar: Los datos personales deben ser procesados al mayor nivel posible de agregación y con el mínimo detalle.

- Ocultar: Los datos personales y sus interrelaciones deben protegerse de forma que no sean visibles para los usuarios.

- Separar: Los datos personales deber ser procesados en entornos separados y distribuidos siempre que sea posible.

- Informar: Los interesados deben ser adecuadamente informados cuando sus datos personales vayan a ser tratados (transparencia).

- Controlar: Los interesados deben poder ejercer sus derechos y conocer en todo momento el procesamiento que se va a realizar de sus datos.

- Cumplir: Debe hacerse cumplir con una política de privacidad compatible con los requerimientos legales.

- Demostrar: Se debe ser capaz de demostrar el cumplimiento de la política de privacidad y de cualquier requerimiento legal aplicable.

15 Big Data. A European survey on the opportunities and riisks of data analytics. Vodafone Institute for Society and Communications TNS. Enero 2016.
16 Privacy by design in big data. An overview of privacy enhancing technologies in the era of big data analytics. Diciembre de 2015 citar la fuente por parte de ISMSForum.

Debido al volumen, la diversidad y la velocidad con la que en ocasiones los datos objeto de tratamiento deben ser procesados en los proyectos de Big data, se introducen varios desafíos adicionales a la hora de definir la estrategia más adecuada de privacidad.

Para hacer frente a dichos desafíos, es necesario conocer los puntos de vista de todas las partes implicadas, así como tener en cuenta la finalidad no sólo del proyecto en cuestión, sino de todas y cada de las fases que conforman la cadena de valor (adquisición, recopilación, análisis, validación, almacenamiento y explotación) de los proyectos de Big Data. De esta manera, será posible extraer los requisitos de privacidad específicos y las medidas de aplicación correspondientes por cada fase.

Sin embargo, es importante matizar que, para poner en práctica un enfoque coherente, aparte de las necesidades de cada fase en particular, hay que tener en cuenta el ciclo de vida completo.

Adicionalmente, no se trata sólo de una técnica u otra, sino más bien de diversas tecnologías gracias a las cuales se podrán cubrir adecuadamente las diferentes estrategias de privacidad definidas en cada una de las fases del ciclo de vida de los proyectos de Big Data.

A continuación, se exponen las distintas técnicas o tecnologías que permitirían cubrir las diferentes estrategias de privacidad:

- **Anonimización**: Serviría para las estrategias de minimizar o agregar.
- **Cifrado**: En el caso de ocultar o separar.
- **Control de Acceso**: Si se trata de informar o controlar.
- **Trazabilidad**: Para las de cumplir o demostrar.

A continuación, se presentan a modo resumen las estrategias de privacidad más adecuadas para cada una de las fases que conforman la cadena de valor de Big Data:

FASE BIG DATA	ESTRATEGIA	IMPLEMENTACIÓN
Adquisición y recolección	Minimizar	• Seleccionar antes de adquirir • EIPD
	Agregar	• Anonimización en la fuente 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

IV.2.- MEDIDAS TÉCNICAS.

Aunque los procesos de anonimización y disociación son clave para respetar la privacidad en los análisis de Big Data, no hay que desdeñar otras medidas técnicas aplicables al desarrollo de cualquier sistema, aunque con particularidades propias para su aplicación a Big Data[17]. Entre estas medidas cabe destacar medidas de cifrado, de control de acceso, medidas de responsabilidad proactiva y medidas de transparencia, consentimiento, monitorización y control.

IV.3 – MEDIDAS PARA MEJORAR LA CONFIANZA.

El nuevo Reglamento General de Protección de Datos de la UE desarrolla en los Artículos 40 a 43 los aspectos principales a considerar para el desarrollo de códigos de conducta, mecanismos de certificación, sellos y etiquetas de protección de datos.

Estas herramientas no sólo sirven para guiar a los responsables y encargados del tratamiento de datos personales en el cumplimiento de los requisitos impuestos por la nueva regulación, sino que, además, contribuyen a demostrar el cumplimiento de las obligaciones legales y refuerzan la confianza de los sujetos interesados y reguladores en las organizaciones que las aplican. Aunque todas buscan ayudar a las organizaciones en el cumplimiento de sus obligaciones legales, hay ligeras diferencias entre ellas que conviene destacar.

Los códigos de conducta están destinados a contribuir a la correcta aplicación del Reglamento teniendo en cuenta las características específicas de los distintos sectores de tratamiento, y serán las asociaciones y otros organismos representativos de categorías de responsables o encargados del tratamiento quienes podrán elaborar códigos de conducta, modificar o ampliarlos con objeto de especificar la aplicación del Reglamento. Por tanto un código de conducta busca asistir a las organizaciones en la aplicación adecuada de la nueva regulación. Los códigos de conducta ya estaban presentes en la Directiva de Protección de Datos, siendo el GT29 el encargado de aprobar aquellos que afectasen a toda la Unión Europea.

En el RGPD, el artículo 40 es la principal fuente de referencia para el establecimiento de códigos de conducta, y en él se establece que serán las autoridades de protección de datos las que alentarán su desarrollo, bien directamente o a través de terceros, y determinarán si proporcionan suficientes garantías.

Una vez aprobados, los códigos de conducta serán publicados en un registro. Adicionalmente, el Artículo 41 establece la posibilidad de acreditar entidades independientes que se encarguen de monitorizar el adecuado cumplimiento del código de conducta por parte de una organización adherida.

En la actualidad existen códigos de conductas en ámbitos técnicos como la gestión de identidades federadas[18], la provisión de servicios en la nube[19], y también en dominios de negocio como la publicidad[20], entre otros. Es de esperar que en próximas fechas aparezcan nuevos códigos de conducta referentes al procesamiento masivo de datos que ayuden a las organizaciones a aplicar la nueva regulación de forma adecuada.

Por su parte, las certificaciones, sellos o etiquetas aparecen en la nueva regulación como un mecanismo aceptado para demostrar el cumplimiento de la regulación. En particular, el artículo 42 establece que se promoverá el establecimiento de mecanismos de certificación y sellos y etiquetas de protección de datos a nivel europeo. Por su parte, el artículo 43 desarrolla los aspectos relacionados con el establecimiento y acreditación de organismos de certificación.

En la actualidad existen diversas certificaciones de privacidad[21], tanto a nivel nacional como internacional. Sin embargo, estas certificaciones varían en objetivos y alcance, y por el momento no han sido acreditadas oficialmente. En Europa, la certificación de privacidad más conocida es EuroPriSe, que ofrece certificaciones para productos y servicios IT que cumplen con la legislación europea de protección de datos.

IV.4.- BUENAS PRÁCTICAS.

Las medidas técnicas y de confianza descritas en esta sección evolucionan constantemente en línea con los vertiginosos avances tecnológicos. Dada esta rápida evolución, hay que tener en cuenta que los procesos realizados pueden no ser definitivos e irreversibles, ya que dependerán del avance de la técnica y de las fuentes de datos conocidas.

La solución óptima debe decidirse caso por caso y puede conllevar la combinación de diversas técnicas, teniendo siempre como objetivo principal el evitar la identificación del interesado sobre todo al usar varias fuentes de información, sean o no accesibles al público. Se considerarán buenas prácticas:

• La metodología de Protección de Datos desde el Diseño (PDdD) debería ser el marco de actuación de todos los procesos descritos, impulsando los análisis de impacto en la protección de datos que deben estar presente en los estudios previos al establecimiento de las medidas técnicas necesarias.

17 https://downloads.cloudsecurityalliance.org/initiatives/bdwg/Expanded_Top_Ten_Big_Data_Security_and_Privacy_Challenges.pdf
18 GÉANT Data Protection Code of Conduct.
19 Data Protection Code of Conduct for Cloud Service Providers.
20 Dictamen 4/2010 relativo al «Código de conducta europeo de la FEDMA sobre la utilización de datos personales en la comercialización directa». GT29. Bruselas 2010.
21 Certification schemes for cloud computing o los que se mencionan en el sitio web de INCIBE.

• Antes de aplicar cualquier técnica de anonimización hay que valorar el uso de la misma en relación a los requisitos previos o contexto y los objetivos o finalidad del proceso de anonimización que buscamos.

• La anonimización realizada se debe ir revisando periódicamente e igualmente evaluando los posibles nuevos riesgos que puedan surgir como consecuencia de diferentes factores, riesgo residual de datos anonimizados, nuevas fuentes de datos a cruzar, nuevas tecnologías, etc.

• Las técnicas de anonimización deben preservar la utilidad de los datos en la medida de lo posible, sin perder de vista el impacto que puede tener la utilización de las mismas, especialmente en el caso de elaboración de perfiles.

• No hay que confundir seudonimización y anonimización. La seudonimización es una técnica que consiste en reemplazar un atributo por otro en un registro, reduce la capacidad de vinculación de un conjunto de datos con la identidad original del interesado, pero sigue permitiendo identificar indirectamente al interesado si no se añaden otras medidas, por tanto, es una medida de seguridad útil a usar como paso intermedio en un proceso de anonimización.

• Se deben establecer medidas adicionales de seguridad en todos los elementos que intervienen en el proceso de la anonimización, como auditorías periódicas de las fuentes de información, de los canales de transmisión de la información, de las localizaciones físicas de las fuentes de información, etc., aplicando los estándares, sellos y buenas prácticas en seguridad y privacidad de la información. Este marco integral debería incluir un procedimiento de detección y notificación de posibles brechas de privacidad que pudieran surgir, como casos de re-identificación.

• Es recomendable la adopción de códigos de conducta en las organizaciones para facilitar la aplicación de la legislación vigente, así como la obtención de certificaciones, sellos o etiquetas que permitan demostrar a terceros su adecuado cumplimiento, de forma que la privacidad se pueda convertir en valor referencial de las mismas.

REFERENCIAS

Para la elaboración del presente Guía de Buenas Prácticas se ha considerado lo dispuesto en las siguientes disposiciones, normativas y documentos:

REGLAMENTO DEL PARLAMENTO EUROPEO Y DEL CONSEJO 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 (RGPD).

Ley Orgánica 15/1999, de 13 de diciembre, de Protección de Datos de Carácter Personal (LOPD).

Real Decreto 1720/2007, de 21 de diciembre, por el que se aprueba el Reglamento de desarrollo de la Ley Orgánica 15/1999, de Protección de Datos (RLOPD).

Dictamen 6/2013, emitido por el Grupo de Trabajo del Artículo 29, sobre protección de datos en la reutilización de la información del sector público.

Dictamen 5/2014, emitido por el Grupo de Trabajo del Artículo 29, sobre técnicas de anonimización.

Orientaciones sobre procedimientos de anonimización de datos personales emitidas por la Agencia Española de Protección de Datos en 2015. (2015)

'Towards a thriving data-driven economy', which sets forth the Commission's strategy on Big Data COM (2014) 442 final.

Opinion 7/2015 Meeting the challenges of Big Data I(European data Protection Supervisor). (2015) https://secure.edps.europa.eu/EDPSWEB/webdav/site/mySite/shared/Documents/Consultation/Opinions/2015/15-11-19_Big_Data_EN.pdf 19/11/2015.

Privacy by design in Big Data: An overview of privacy enhancing technologies in the era of Big Data analytics (European Union Agency For Network And Information Security). https://www.enisa.europa.eu/media/news-items/privacy-by-design-in-big-data-an-overview-of-privacy-enhancing-technologies-in-the-era-of-big-data-analytics 17/12/2015.

"Opinion 03/2013 on purpose limitation". (2 April 2013).

Agencia de los Derechos Fundamentales de la Unión Europea y Consejo de Europa (2014). Manual de legislación europea en materia de la protección de datos. Luxemburgo: Oficina de Publicaciones de la Unión Europea.

APARICIO SALOM, J. (2013). Estudio sobre la Protección de Datos. Aranzadi. Capítulo VII. Apartado XII.I.

European Data Protection Supervisor: Opinion 7/2015 Meeting the challenges of Big Data (2015).

Federal Trade Commission: Big Data a Tool for Inclusion or Exclusion, https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf January 2016.

Information Commissioner`s Office: Big Data and data protection.

Mayer-Schönberger, V. y Cukier, K. (2013). Big Data. La revolución de los datos masivos. Turner.

Recomendación CM/ Rec (2010) 13, del Comité de Ministros a los Estados miembros sobre la protección de las personas con respecto al tratamiento automatizado de datos de carácter personal en el contexto de la creación de perfiles.

SEMPERE SAMANIEGO, F. J. (2013). Comentarios prácticos a la Propuesta de Reglamento de Protección de Datos de la Unión Europea. Licencia Creative Commons (CC BY-NC-SA).

CNIL. Metodología para la gestión de riesgos para la Privacidad.

El enfoque de ENISA (European Union Agency for Network and Information Security) para la Gestión de riesgos.

Guía para una evaluación de impacto en la protección de datos personales. AEPD. 2014.

Étude d'Impact sur la vie Privée (EIVP). Comment mener une EIVP. CNIL francesa. Junio de 2015.

Benefit-Risk Analysis for Big Data Projects. Future of Privacy Forum. Septiembre 2014.

The 7 Foundational Principles. Implementation and Mapping of Fair Information Practices. (Ann Cavoukian, Ph.D.-Information & Privacy Commissioner. Ontario, Canada).

Privacy By Design- Protecting privacy in the age of analytics (Deloitte).

Privacy and Data Protection by Design – from policy to engineering (ENISA – European Union Agency for Network and Information Security).

32nd International Conference of Data Protection and Privacy Commissioners. Privacy by design resolution.

Más información en:

AGENCIA ESPAÑOLA DE PROTECCIÓN DE DATOS

www.agpd.es

isms forum spain

www.ismsforum.es