novática

Revista de la Asociación de Técnicos de Informática

Novática, revista fundada en 1975 y decana de la prensa informática española, es el órgano oficial de expresión y formación continua de **ATI** (Asociación de Técnicos de Informática), organización que edita también la revista **REICIS** (Revista Española de Innovación, Calidad e Ingeniería del Software).

<http://www.ati.es/novatica/>
<http://www.ati.es/reicis/>

ATI es miembro fundador de **CEPIS** (*Council of European Professional Informatics Societies*), representa a España en **IFIP** (*International Federation for Information Processing*) y es miembro de **CLEI** (*Centro Latinoamericano de Estudios de Informática*) y de **CECUA** (*Confederation of European Computer User Associations*). Asimismo tiene un acuerdo de colaboración con **ACM** (*Association for Computing Machinery*) y colabora con diversas asociaciones informáticas españolas.

Coordinación Editorial, Redacción Central y Redacción ATI Madrid
Gutierre de Cetina 24, 28017 Madrid • Tlfn.914029391 <novatica@ati.es>
Administración y Redacción ATI Cataluña
Calle Avila 50, 3a planta, local 9, 08005 Barcelona
Tlfn.934125235 <secregen@ati.es>
Redacción ATI Andalucía <secreand@ati.es>
Redacción ATI Galicia <secregal@ati.es>
Suscripción y Ventas <novatica.subscripciones@atinet.es>
Publicidad Gutierre de Cetina 24, 28017 Madrid
Tlnf.914029391 <novatica@ati.es>
Imprenta: Impresión Offset Derra S.L., Llull 41, 08005 Barcelona.
Depósito legal: B 15.154-1975 -- ISSN: 0211-2124; CODEN NOVAEC
Portada: "La decisión" - Concha Arias Pérez / © ATI
Diseño: Fernando Agresta / © ATI 2003

sumario

Nº 237, julio-octubre 2016, año XLII

Monografía del próximo número: "Seguridad digital"

Privacidad en la analítica masiva de datos

José María del Álamo Ramiro[1], Esmeralda Saracibar Serradilla[2], Emilio Aced Félez[3]
[1]Departamento de Ingeniería de Sistemas Telemáticos, Escuela Técnica Superior de Ingeniería de Telecomunicación, Universidad Politécnica de Madrid; [2]ECIX Group; [3]Unidad de Evaluación y Estudios Tecnológicos, Agencia Española de Protección de Datos

<jm.delalamo@upm.es>, <esmeralda.saracibar@ecixgroup.com>

En memoria de nuestro compañero Emilio Aced Félez que falleció mientras este artículo estaba en proceso de publicación.

Resumen: *La analítica masiva de datos promete revelar relaciones entre datos hasta ahora ocultas, predecir tendencias generales y ofrecer nuevo conocimiento inferido mediante la recolección de enormes y complejos conjuntos de datos y la aplicación de novedosos algoritmos de análisis. Este avance tecnológico trae consigo grandes posibilidades para organizaciones de distintos tipos y tamaños, que ven en él una posibilidad para mejorar sus procesos, resultando tanto en beneficios sociales como económicos. Sin embargo, los avances tecnológicos normalmente llevan asociados nuevos desafíos para las sociedades que los adoptan, y la analítica masiva de datos plantea notables riesgos para la privacidad de las personas por su capacidad para revelar información que se creía privada. En este artículo se ofrece una panorámica general sobre los principales riesgos que la analítica masiva de datos entraña para la privacidad y la protección de datos personales, y se ofrecen distintos enfoques jurídicos, organizativos y técnicos que las organizaciones pueden adoptar para sacar el máximo provecho de los datos de que disponen a la vez que respetan los derechos individuales.*

Palabras clave: *Análisis de datos,* Big Data*, GDPR, inferencia, privacidad, procesamiento, protección de datos personales, recolección, regulación.*

1. Introducción

La analítica masiva de datos, o *Big Data*, es el nombre que se le da al conjunto de tecnologías, algoritmos y sistemas empleados para recolectar y almacenar datos y extraer información de valor de ellos mediante sistemas analíticos avanzados, todo ello llevado a cabo sobre ingentes y variados volúmenes de datos [1]. Es fácil ver que la capacidad de tratar grandes colecciones de datos (muchos de ellos personales), y desarrollar inferencias y detectar correlaciones lleva aparejadas enormes posibilidades de desarrollo y a la vez retos importantes para la privacidad y la protección de datos personales a los que hay que hacer frente.

En efecto, la analítica masiva de datos puede producir importantes beneficios sociales en muchos campos como la investigación científica y médica, la mejora de la asistencia sanitaria, la detección del fraude, o la mejora en la asignación de recursos por parte de organismos públicos, entre otros. Y, por supuesto, beneficios económicos a las organizaciones que la empleen para conocer mejor a sus clientes y usuarios, y establecer sus estrategias comerciales de acuerdo con ese conocimiento.

Por ejemplo, en el sector de la distribución permite obtener conocimiento para anticiparse al consumidor evitando situaciones de desabastecimiento de productos y falta de suministro; en el sector sanitario permite reducir el tiempo de ingreso hospitalario o predecir futuras enfermedades y riesgos sanitarios; la investigación médica se ha beneficiado del desarrollo de herramientas que permiten el procesamiento de la enorme cantidad de datos generados en relación con el genoma humano; y, se prevé un uso extensivo de estas tecnologías para la gestión y prevención de las incidencias de tráfico y excesos de contaminación en las ciudades.

Por otro lado, también surgen dudas y preocupaciones sobre posibles usos que, o bien no sean lícitos por realizarse sin respaldo legal para ello, o bien generen incertidumbre y desconfianza en la sociedad, derivados del hecho de que estas tecnologías no sólo

Autores

José María del Álamo Ramiro es doctor e ingeniero de telecomunicación, y actualmente es profesor en la Universidad Politécnica de Madrid (UPM) donde enseña distintos aspectos de la ingeniería de sistemas y servicios telemáticos a nivel de grado, máster y doctorado. Investiga en gestión de la identidad, privacidad y confianza desde 2005, colaborando con empresas en sectores como las telecomunicaciones (Telefónica, Ericcson), banca y servicios financieros (Santander), redes sociales (Vodafone, Twitter) o transporte y logística (Transfesa – Deutsche Bahn AG), participando en consorcios y proyectos nacionales e internacionales, y representando a la UPM ante organismos de estandarización como OASIS o W3C. Organiza el *IEEE International Workshop on Privacy Engineering* en San José, California, y es autor de más de 40 publicaciones y 5 patentes internacionales en la temática.

Esmeralda Saracibar Serradilla es socia del área de *Governance, Risk & Compliance* de Ecix. En su trayectoria profesional cuenta con un extenso recorrido tras 15 años dedicada a la asesoría jurídica de empresas en asuntos de Derecho y nuevas tecnologías, privacidad de datos, *Compliance*, *Big Data*, *Cloud Computing*, *Corporate Compliance*, Comercio Electrónico, etc. Es licenciada en Derecho por la Universidad de Deusto (2001), Máster Derecho de las Telecomunicaciones ICAI-ICADE (2002) y Máster en Dirección de Seguridad Global de la Universidad Europea Madrid (2004). Ostenta el certificado de ISMS Forum como analista de riesgos de Seguridad de la Información (2009) y el CDPP *Certified Data Privacy Professiona*l del Data Protection Institute (2011). Forma parte del capítulo español del CSA *Cloud Security Alliance* y del Comité Operativo del DPI y participa en el Grupo de Trabajo del DPI para la elaboración de una guía de buenas prácticas ante iniciativas de *Big Data*. Participa como ponente en cursos especializados en instituciones de reconocido prestigio como el Instituto de Empresa en el que imparte el Programa Especializado de *Compliance*. Ha sido ponente en el Master Ejecutivo de Dirección de Seguridad Global de la Universidad Europea de Madrid (2003-2006) y en el curso especializado en Seguridad de la Información del Ilustre Colegio de Abogados de Madrid). Ha publicado como co-autora las obras: *Compliance. Cumplimiento normativo y seguridad en la empresa* y *Protección de datos personales* (Thomson - Aranzadi). Es colaboradora habitual en revistas y prensa especializada: Diario Jurídico, El Derecho, Legal Today, El Economista, Economist & Jurist, etc.

Emilio Aced Félez fue Jefe de Área en la Agencia Española de Protección de Datos (AEPD). Durante los últimos años desempeñó su actividad en el ámbito de la protección de datos personales tanto en la AEPD como Subdirector General de Inspección y Adjunto al Director y encargado de las relaciones internacionales de la AEPD; como en la Agencia de Protección de Datos de la Comunidad de Madrid en la que fue Subdirector General de Inspección y Tutela de Derechos y Subdirector General de Registro de Ficheros y Consultoría. Representó a la AEPD en el Grupo de Trabajo de Autoridades Europeas de Protección de Datos (Grupo de Trabajo del Artículo 29) y el Comité Consultivo del Convenio 108 del Consejo de Europa y fue miembro de las Autoridades Comunes de Control de Schengen, Sistema de Información Aduanero y Europol, de la cual fue Vicepresidente y Presidente.

> “ Las iniciativas de *Big Data* conllevan la adopción de soluciones de cumplimiento en materia de protección de datos, desde un punto de vista jurídico, organizativo y técnico, con el fin de llevar a cabo un adecuado respeto de los derechos de los individuos y una idónea gestión de riesgos de incumplimiento por parte de las organizaciones ”

son capaces de describir el pasado y predecir el futuro con cierto nivel de incertidumbre, sino incluso de condicionarlo con las decisiones adoptadas [2].

Hoy en día son conocidos casos desafortunados donde el uso de técnicas de *Big Data* ha resultado en la violación de los derechos de privacidad y protección de datos de individuos.

Por ejemplo, el empleo de herramientas de analítica de datos para la identificación automática de clientes a término de su periodo de gestación y el posterior envío de material publicitario relacionado desencadenó una oleada de críticas al desvelar el embarazo de una menor de edad de forma imprevista por la afectada [3].

También han sido muy criticados los experimentos de condicionamiento inducido a usuarios de redes sociales a partir de la recolección de ingentes cantidades de datos generados por ellos y sus contactos, su análisis para la identificación de estados de ánimo, y su publicación de forma selectiva para condicionar estados de ánimo particulares en los receptores [4].

Este artículo aborda la problemática, y presenta algunos enfoques de solución, para mantener el equilibrio entre un uso legítimo y adecuado de las tecnologías de *Big Data* por parte de las organizaciones y el respeto a la privacidad y la protección de datos de los individuos.

Para ello, la siguiente sección presenta los riesgos más destacados asociados al análisis masivo de datos y, a continuación, se introducen distintos enfoques de solución, considerando aspectos legales, organizativos y técnicos. Por último, y dada la necesidad de abordar soluciones integrales ante un problema tan complejo, se ofrecen referencias para profundizar en aspectos particulares.

2. Principales riesgos para la privacidad

Los proyectos de *Big Data* siempre se concretarán en escenarios complejos desde el punto de vista de su interrelación con el derecho fundamental a la protección de datos personales. Se trata de procesamientos masivos que afectan a millones de personas, para finalidades que puede que no estuvieran previstas cuando se recogieron los datos y cuyos resultados, en muchos casos y si no se utilizan adecuadamente, pueden tener efectos muy adversos sobre la vida de las personas.

De hecho, el nuevo Reglamento 2016/679 (UE), General de Protección de Datos (RGPD) [5][1], aunque no habla en ningún momento de *Big Data* como tal, sí refleja la preocupación del legislador por dos de los resultados habituales de los tratamientos de *Big Data*: la utilización de sus resultados para la toma de decisiones automatizadas sobre las personas y la utilización de sus resultados para la elaboración de perfiles, estableciendo previsiones sobre la necesidad de informar al ciudadano y entregar esta información cuando se ejerce el derecho de acceso. Por lo tanto, existen una serie de riesgos genéricos que los tratamientos de *Big Data* tienen para la protección de datos (independientemente de los específicos de cada proyecto concreto).

La falta de trasparencia u opacidad de los mismos es uno de los más importantes. En efecto, los tratamientos de *Big Data* se llevan a cabo en la mayoría de las ocasiones con los datos existentes en las organizaciones (complementados con datos de fuentes externas en muchos casos) sin que los afectados tengan ningún conocimiento de ello y no puedan expresar, como mínimo, su oposición a la utilización de los mismos.

El otro riesgo importante, conectado de alguna manera con el anterior, es la utilización de los datos con fines incompatibles respecto de los que se indicaron cuando se recogieron los datos y sin una legitimación adecuada para hacerlo.

Otra consideración muy importante es la calidad de los datos que se usan en los análisis. En muchas ocasiones no existe ninguna garantía de que los mismos tengan una mínima calidad y, por tanto, las predicciones que se realicen pueden ser completamente erróneas, que es otro de los riesgos: el creer a pies juntillas que los resultados son algo más que correlaciones estadísticas entre variables que pueden no reflejar la realidad.

Finalmente, hay que mencionar el potencial discriminatorio que tiene el *Big Data* y el riesgo de que se nos encasille en diversos perfiles que tengan efectos muy relevantes sobre nuestra vida en el acceso a servicios como el crédito, los seguros o la salud.

Además, también es necesario tener en cuenta la temporalidad de las predicciones. Una predicción o una inferencia que se ha hecho hoy no necesariamente sigue siendo válida un año después, entre otras cosas porque la utilización de la predicción altera la realidad y puede transformarla en un plazo no demasiado grande de tiempo.

Teniendo en cuenta estos riesgos, las organizaciones pueden adoptar una serie de medidas legales, organizativas y técnicas que contribuyan a mitigarlos o evitarlos completamente, y que se introducen en el siguiente apartado.

3. Enfoques de solución

Las iniciativas de *Big Data*, si bien aportan valor a las organizaciones, conllevan la adopción de soluciones de cumplimiento en materia de protección de datos, desde un punto de vista jurídico, organizativo y técnico, con el fin de llevar a cabo un adecuado respeto de los derechos de los individuos y una idónea gestión de riesgos de incumplimiento por parte de las organizaciones.

En general, en primer lugar, es necesario valorar si la iniciativa requiere necesariamente el tratamiento de datos personales, es decir, asociados a una persona concreta, identificable o que pudiera ser identificada. En el caso de que sea necesario o se tenga previsto utilizar datos personales, habrá de tenerse en cuenta la aplicación de la normativa de protección de datos.

Habida cuenta del contexto en el que nos encontramos, habrán de observarse tanto la normativa local actual (Ley Orgánica de Protección de Datos –LOPD, y el Regla-

> “ La disociación es el proceso por el cual se hacen anónimos distintos datos personales, es decir, se generan datos que no permiten la identificación del afectado o interesado con el que estaban relacionados ”

mento que la desarrolla - RLOPD) como la reciente normativa europea (RGPD), aplicable a partir de mayo de 2018.

Por el contrario, si la iniciativa no requiere el uso de datos personales, se reducirán de forma considerable los riesgos asociados. Por ello, a continuación describimos por separado cada uno de estos casos.

3.1. Análisis que no requiere datos personales

En los procesos de análisis masivos de datos que no requieren el uso de datos personales se puede establecer un proceso preliminar denominado de disociación o anonimización.

La disociación es el proceso por el cual se hacen anónimos distintos datos personales, es decir, se generan datos que no permiten la identificación del afectado o interesado con el que estaban relacionados. La disociación es un proceso relevante para la gestión de la privacidad en los procesos de *Big Data* ya que permite anonimizar conjuntos de datos de forma que éstos dejen de contener información considerada personal, y pueda ser tratada para generar conocimiento para el negocio sin las restricciones asociadas a los datos personales. Dicho de otra forma, al disociar la información que una organización maneja se reducen los riesgos de infringir la normativa de protección de datos, puesto que la organización reduce la cantidad de información personal que debe gestionar.

Existen distintas técnicas que permiten llevar a cabo la disociación de un conjunto de datos, que básicamente se pueden categorizar en técnicas de supresión, generalización y aleatorización.

Las primeras son las más burdas, y se centran en suprimir los registros más comprometidos. Las técnicas de generalización buscan diluir valores particulares de un dato de forma que ese valor sea compartido por un número suficiente de afectados para evitar que se pueda singularizar a un afectado (y el resto de sus datos asociados) conociendo exclusivamente ese valor.

Por ejemplo, en lugar de usar cinco dígitos para representar el código postal se pueden utilizar los cuatro más significativos, tres o incluso sólo dos, con lo que el número de afectados que comparten el mismo valor para este atributo crece, reduciendo el riesgo de identificar a una persona en particular al conocer su código postal. Por su parte, las técnicas de aleatorización modifican el valor del dato para un individuo pero buscando conservar las propiedades observables para el conjunto de datos.

Como cabe esperar cada técnica presenta sus ventajas e inconvenientes. Primero, su aplicación lleva asociada una pérdida de información distinta, lo que finalmente reduce la utilidad, y por lo tanto el valor, del conjunto de datos resultante. Por otro lado, los resultados ofrecen distintas garantías frente a la re-identificación, por ejemplo, al combinarse con otros conjuntos de datos.

Es importante destacar que para garantizar la irreversibilidad se habrán de considerar tanto (1) las potenciales fuentes de información externas disponibles en los diferentes medios, especialmente en internet, como (2) la tecnología aplicable, no solo por parte del responsable del tratamiento sino por cualquier otra persona. De lo contrario, se corre el riesgo de que, como ha ocurrido en el pasado [6][7], conjuntos de datos considerados anónimos realmente no lo sean.

Las organizaciones deben evaluar los beneficios e inconvenientes de cada técnica y decantarse por aquella que en cada caso resulte óptima [GTA29].

Para llevar a cabo un correcto proceso de disociación las organizaciones deberían implantar una serie de medidas adicionales:

- **Definición de una política de disociación**: resulta aconsejable definir una política de disociación que se encuentre documentada y actualizada, de manera que refleje de forma justificada las actuaciones a llevar a cabo para proteger la privacidad de los interesados y se encuentre accesible al personal implicado en el tratamiento de datos disociados, así como un protocolo de actuación del proceso de disociación.
- **Uso de sellos de tiempo**: adicionalmente se habrá de tener en cuenta la posibilidad de utilizar en el proceso de anonimización algoritmos de sello de tiempo, con el fin de garantizar la fecha y hora en la que se realizó, o incluso algoritmos de firma electrónica para garantizar la identidad electrónica de quien ha realizado el proceso.
- **Establecimiento de garantías jurídicas adicionales**: La política de disociación, 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 (1) acuerdos de confidencialidad y cláusulas contractuales que garanticen la privacidad de la información incluso cuando haya brechas de re-identificación; (2) 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 (3) auditorías de uso de la información anonimizada.

Por último, sería recomendable 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 viabilidad de las técnicas de anonimización propuestas y del procedimiento de disociación.

3.2. Análisis que utiliza datos personales

Aquellas organizaciones que desarrollan análisis masivos de datos en los que intervienen datos personales deberán llevar a cabo procedimientos conducentes a mitigar o evitar los riesgos para la privacidad de los afectados o interesados, y con ello reducir sus propios riesgos de incumplimiento de la legislación vigente.

Para ello, como primera medida, la organización debería establecer una política de protección de datos desde el diseño y por defecto [9], para que desde las etapas iniciales de análisis de viabilidad de un proyecto se tenga en cuenta la necesidad de que el desarrollo respete los principios de protección de datos y lleve embebidos todos los requerimientos en esta materia.

De hecho, desde hace años, son muchas las voces que reclaman que cualquier proceso de desarrollo de servicios o sistemas debe tener en cuenta los aspectos de privacidad desde sus fases iniciales y a lo largo de todo el ciclo de vida [10]. Por ejemplo, el RGPD indica que “*…el responsable del tratamiento debe […] aplicar medidas que cumplan en particular los principios de protección de datos desde el diseño…*”.

> “ Como primera medida, la organización debería establecer una política de protección de datos desde el diseño y por defecto para las etapas iniciales de análisis de viabilidad de un proyecto ”

Si bien hay acuerdo en que las consideraciones de privacidad no deben realizarse una vez que se ha concluido un desarrollo, el proceso concreto a seguir para desarrollar un servicio o sistema que respete los principios básicos de privacidad no está tan claro.

En la actualidad, existen múltiples metodologías que ofrecen soluciones parciales para distintos dominios de negocio y/o contextos [11], varias técnicas que abordan problemas de privacidad particulares, e incluso herramientas de soporte. Aun así, la disciplina que permita identificar y abordar de forma sistemática y óptima los desafíos de privacidad a los que un ingeniero se enfrenta está todavía en desarrollo [12], aunque con resultados prometedores.

Por ejemplo, Apple ha anunciado recientemente la introducción de herramientas novedosas para conjugar sus necesidades de explotación de datos con el respeto a los derechos de privacidad y protección de datos [13].

La política de protección de datos debe ser vinculante y aplicable, con el fin de asegurar y evidenciar la verificación de cumplimento de las exigencias legales y contemplando, como mínimo, los siguientes aspectos:

- **Limitación del tratamiento** de datos personales al mínimo imprescindible, y a las finalidades informadas y, en su caso, consentidas.
- **Licitud del tratamiento** en cuanto a que su origen sea legítimo, su tratamiento sea proporcional y no excesivo.
- **Transparencia**, en cuanto a la actitud del personal en la asunción del compromiso de actuar conforme a los objetivos de cumplimiento que previamente se hayan establecido.
- **Formación al personal** en el contenido de la misma.
- **Mecanismos de evaluación** de su fiabilidad y efectividad.
- **Auditorías** externas y/o internas de su cumplimiento.

De cara a observar algunos de estos aspectos se deberán introducir cláusulas de consentimiento informado al interesado, transparentes, expresas, precisas e inequívocas sobre las finalidades previstas, tales como, la elaboración de perfiles y modelos predictivos basados en la información del individuo, sus patrones de conducta, hábitos y preferencias, así como, en un futuro, sobre los criterios lógicos, o algoritmos utilizados para la creación de su perfil y de las consecuencias que la creación de esos perfiles tendrán para los mismos y, a partir de la cual, se obtenga el consentimiento libre, específico, informado e inequívoco, de los afectados.

Además, deberán tenerse en cuenta aquellos supuestos en los que los datos se hayan obtenido de fuentes distintas del interesado, por ejemplo, fuentes de acceso público o de bases de datos comercializadas por terceros.

Especial relevancia adquieren los supuestos en los que las iniciativas de *Big Data* sean implantadas para decidir, por ejemplo, sobre el perfeccionamiento, finalización y/o renovación de la relación contractual suscrita, basándose únicamente en un tratamiento de datos destinados a evaluar determinados aspectos de su personalidad, a partir de valoraciones automatizadas, donde será necesario informar al cliente/usuario sobre la posibilidad de impugnar dichas valoraciones.

Otro aspecto muy importante, que se puede considerar como una herramienta para implantar las políticas de protección de datos desde el diseño, son las evaluaciones de impacto en la protección de datos (EIPD, o PIA —*Privacy Impact Assessment*— por sus siglas en inglés).

Una evaluación de impacto en la protección de datos personales es, básicamente, un ejercicio de análisis de riesgos para detectar los que puede entrañar el nuevo producto o servicio para las personas cuyos datos trata.

Concretamente el RGPD, establece la necesidad de realizar la EIPD siempre que se lleven a cabo elaboraciones de perfiles, en especial 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. Consecuencia de lo anterior, las organizaciones habrán de realizar una EIDP ante iniciativas de *Big Data* que involucren datos personales.

Para ayudar a las organizaciones a llevar a cabo estas evaluaciones de impacto, la Agencia Española de Protección de Datos (AEPD) publicó ya en 2014 una *Guía para una evaluación de impacto en la protección de datos* [14][2].

La gran ventaja derivada de la realización de una EIPD es que permite identificar los posibles riesgos y corregirlos anticipadamente, antes de que el producto esté en el mercado y pueda lesionar los derechos de las personas y también tener un coste importante para la reputación de la organización.

En función de los resultados del análisis de riesgos (riesgos respecto de los derechos fundamentales de los interesados), habrá que implantar las medidas técnicas que se consideren idóneas para asegurar la privacidad. Entre las medidas técnicas que pueden contribuir a mitigar y corregir estos riesgos se encuentran las técnicas de disociación ya mencionadas en el apartado anterior, pero también otras como las técnicas de cifrado u otras para protección de la información, correspondientes al nivel de seguridad de la tipología de datos personales que sean objeto del tratamiento. En este sentido, al tratarse de la creación de perfiles basados en las características del individuo, serán de aplicación, como mínimo, las medidas de seguridad de nivel medio.

Las técnicas de cifrado buscan garantizar la confidencialidad de los datos personales, al impedir su acceso por quienes no dispongan de las claves de descifrado.

Estas técnicas pueden aplicarse cuando los datos están almacenados (por ejemplo, en un fichero o base de datos) y también mientras están en tránsito (por ejemplo, mediante uso de protocolos de comunicación seguros). Los sistemas comerciales de procesamiento masivo de datos suelen integrar este tipo de soluciones para minimizar el riesgo de revelación de datos, aunque de momento trabajan con los datos descifrados.

Para mitigar el riesgo de fuga de datos durante el procesamiento se están desarrollando soluciones que son capaces de realizar operaciones sobre datos cifrados (englobadas en el denominado cifrado homomórfico), lo que sin duda contribuirá a reducir más aún los riesgos.

Cabe mencionar que existen otras medidas técnicas que el responsable de la protección de datos debe considerar como un adecuado control de acceso a los datos, la trazabilidad de las operaciones realizadas, la implementación de procesos automáticos de monitorización y depuración de los datos,

> “ Otro aspecto que deberían tener en cuenta todas aquellas organizaciones que llevan a cabo este tipo de tratamientos es la necesidad de contar con un equipo adecuado para ello ”

la definición de los periodos de retención y la aplicación automática de mecanismos de borrado una vez transcurridos éstos, etc.

Otro aspecto que deberían tener en cuenta todas aquellas organizaciones que llevan a cabo este tipo de tratamientos es la necesidad de contar con un equipo adecuado para ello.

En primer lugar, se debería nombrar un delegado de protección de datos para contar con asesoramiento de calidad sobre privacidad y protección de datos y para que supervise que todas las operaciones se llevan a cabo de manera respetuosa con la normativa. Este nombramiento será obligatorio con el RGPD cuando la actividad principal del negocio consista en operaciones de transformación que requieran de un seguimiento regular o sistemático de los interesados a gran escala.

Además, se requiere la presencia en el equipo de diseño y desarrollo de personas con conocimientos y experiencia en privacidad y protección de datos para asesorar y validar los resultados según se vaya avanzando en las fases del proyecto y en todo el ciclo de vida del mismo.

Por último, no hay que olvidar que los proyectos de análisis masivos de datos están sujetos a las mismas obligaciones que el resto de proyectos que tratan datos personales, y que no pueden dejarse de lado, tales como:

- Notificación del fichero a la AEPD cuya finalidad sea la implantación de iniciativas de *Big Data* o declarar esta finalidad en aquellos ya declarados que se encuentren afectados por dichos fines.
- Disponer de protocolos de actuación de ejercicios de derechos de acceso, rectificación, cancelación u oposición al tratamiento y/o de revocación del consentimiento para la gestión de las solicitudes de ejercicio de tales derechos derivadas de esta tipología de iniciativas. La posibilidad de revocación del consentimiento requiere de controles organizativos y técnicos para hacerlo efectivo.
- Cuando la iniciativa cuente con la participación de colaboradores externos o prestadores de servicios, se deberá regular contractualmente el acceso a los datos, con referencia a las medidas de seguridad que colaborador deberá observar.

Todas estas herramientas y metodologías que se han descrito tienen también la función de poner de manifiesto un elemento crucial del nuevo RGPD: la responsabilidad activa de las organizaciones en la demostración del cumplimiento de la norma, implantando los elementos necesarios para demostrar que han puesto en marcha políticas adecuadas de protección de datos.

4. Conclusiones

Este artículo ha presentado los principales riesgos para la protección de datos personales derivados de procesos de análisis masivos de datos, presentando un conjunto de medidas que pueden ser útiles para las organizaciones que llevan a cabo estos procesos.

Estas medidas son tanto de tipo jurídico, como organizativo y también técnico, basadas en los principios de protección de datos desde el diseño y por defecto, y soportadas por una evaluación de impacto para la protección de datos.

Con el objeto de profundizar en estos temas y de ofrecer más detalle que puedan ayudar a las organizaciones en el respeto de los principios de protección de datos a la vez que desarrollan sus procesos de negocio con soporte de tecnologías de *Big Data,* el *Data Privacy Institute* de la Asociación Española para el Fomento de la seguridad de la información, ISMS Forum Spain <http://www.ismsforum.es> ha promovido y está próximo a publicar una guía de protección de datos para proyectos de *Big Data* en el cual se abordan con más detalle los aspectos que deben tenerse en cuenta ante iniciativas de esta tipología y que puede ser un documento útil y de interés para muchas empresas. Invitamos a todos los lectores interesados en la temática a consultar este documento.

Agradecimientos

Este artículo ha contado con la colaboración de la Asociación Española para el Fomento de la seguridad de la información, ISMS Forum Spain, a través de su iniciativa *Data Privacy Institute*, en la que profesionales de la seguridad y la protección de datos, representantes de empresas, universidades e instituciones, han trabajado conjuntamente para desarrollar la primera guía en el ámbito nacional de protección de datos para proyectos de *Big Data*.

Referencias

[1] M.A. Beyer, D. Laney. *The importance of ‘big data’: a definition*, Gartner, pp. 2014-2018, 2012.

[2] The White House. Big Data and Privacy: a technological perspective. *Executive Office of the President, President’s Council of Advisors on Science and Technology*, 2014.

[3] C. Duhigg. How companies learn your secrets. *The New York Times 16, 2012*. <http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html>.

[4] A.D. Kramer, J.E. Guillory, J.T. Hancock. Experimental evidence of massive-scale emotional contagion through social networks, *Proceedings of the National Academy of Sciences 111.24*, pp. 8788-8790, 2014.

[5] EUR-Lex. Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento general de protección de datos). *Document 32016R0679*. <http://data.europa.eu/eli/reg/2016/679/oj>.

[6] A. Narayanan, V. Shmatikov. *How to Break Anonymity of the Netflix Prize Dataset*, The University of Texas at Austin, 2006.

[7] L. Sweeny. k-anonymity: A Model for Protecting Privacy., International Journal on Uncertainty, Fuzziness and Knowledge based Systems 10, pp: 557- 570, 2012.

[8] Comisión Europea. Grupo de Trabajo sobre Protección de Datos del Artículo 29, *Dictamen 05/2014 sobre técnicas de anonimización*, 10 de abril de 2014. <http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp216_es.pdf>.

[9] G. D’Acquisto, J. Domingo-Ferrer, P. Kikiras, V. Torra, Y.A. de Montjoye, A. Bourka. *Privacy by design in big data: an overview of privacy enhancing technologies in the era of big data analytics*. ENISA, diciembre 2015. <https://www.enisa.europa.eu/publications/big-data-protection>.

[10] 32nd International Conference of Data Protection and Privacy Commissioners. *Resolution on Privacy by Design*: Jerusalem, 2010.

[11] N. Notario, A. Crespo, Y.S. Martín-García, J.M. Del Álamo, D. Le Métayer, T. Antignac, A. Kung, I. Kroener, D. Wright. PRIPARE: Integrating Privacy Best Practices into a Privacy Engineering Methodology, *Actas de IWPE15 en IEEE Security and Privacy Workshops (SPW)*, pp. 151-158, San Jose (CA), 21 May 2015.

[12] S. Gürses, J.M. Del Álamo. Privacy Engineering: Shaping an Emerging Field of Research and Practice. *IEEE Security and Privacy, vol. 14, no. 2*, pp. 40-46, 2016.

[13] K. Conger, N. Lomas. What Apple’s differential privacy means for your data and the future of machine learning. *TechCrunch, junio 2016*. <https://techcrunch.com/2016/06/14/differential-privacy/>.

[14] Agencia Española de Protección de Datos. *Guía para una evaluación de impacto en la protección de datos personales*, 2014.

Notas

[1] Toda la legislación de la Unión Europea está accesible en Eur Lex <www.eurlex.eu>.

[2] Se pueden consultar y descargar éste y otros documentos en el apartado *Publicaciones del Canal Resoluciones y Documentos* del sitio web de la AEPD <www.agpd.es>.