DIPUTACIÓ DE VALÈNCIA

Protecció de Dades i Seguretat de la Informació

Cuadernos DivalData

Cuadernos dirigidos a delegados,
responsables y especialistas en protección
de datos personales

Cuaderno nº 18 | Diciembre 2021

LA SEUDONIMIZACIÓN Y ANONIMIZACIÓN DE DATOS PERSONALES

Í N D I C E

LA SEUDONIMIZACIÓN Y ANONIMIZACIÓN DE DATOS PERSONALES

	Página
INTRODUCCIÓN	2
LA SEUDONIMIZACIÓN NO ES LO MISMO QUE LA ANONIMIZACIÓN	4
ASPECTOS A TENER EN CUENTA PARA ABORDAR LA ELIMINACIÓN DE LA IDENTIFICACIÓN EN LA ANONIMIZACIÓN	5
TECNICAS DE ANONIMIZACIÓN: CLAVES	6
GARANTÍAS DE LAS TECNICAS DE ANONIMIZACIÓN	9
MATERIAL COMPLEMENTARIO Y NOTICIAS	10

Os invitamos a trasladarnos aquellas temáticas que resulten de vuestro interés para los próximos boletines informativos. Estas peticiones deberán dirigirse a:

Diputación de Valencia

Dpto. de Protección de Datos y Seguridad de la Información

Pl. de Manises, 4 46003 Valencia

email: dpdsi@dival.es

SUSCRIPCIONES

Si deseas suscribirte a nuestra publicación accede al siguiente

enlace

"TRANSFORMAR UN CONJUNTO DE DATOS PERSONALES EN INFORMACIÓN ANÓNIMA O SEUDONIMIZADA EXIGE REALIZAR UN TRATAMIENTO SOBRE DICHOS DATOS PERSONALES. EL TRATAMIENTO DE ANONIMIZACIÓN GENERA UN ÚNICO Y NUEVO CONJUNTO DE DATOS, MIENTRAS QUE EL TRATAMIENTO DE SEUDONIMIZACIÓN GENERA DOS NUEVOS CONJUNTOS DE DATOS: LA INFORMACIÓN SEUDONIMIZADA Y LA INFORMACIÓN ADICIONAL QUE PERMITE REVERTIR LA ANONIMIZACIÓN."

INTRODUCCIÓN

La información anónima es un conjunto de datos que no guarda relación con una persona física identificada o identificable (Considerando 26 del RGPD), en tanto que la información seudonimizada es un conjunto de datos que no puede atribuirse a un interesado sin utilizar información adicional, requiere que dicha información adicional figure por separado y, además, esté sujeta a medidas técnicas y organizativas destinadas a garantizar que los datos personales no se atribuyan a una persona física identificada o identificable (Artículo 4.5).

Transformar un conjunto de datos personales en información anónima o seudonimizada exige realizar un tratamiento sobre dichos datos personales. El tratamiento de anonimización genera un único y nuevo conjunto de datos, mientras que el tratamiento de seudonimización genera dos nuevos conjuntos de datos: la información seudonimizada y la información adicional que permite revertir la anonimización.

El conjunto de datos anonimizados no está bajo el ámbito de aplicación del Reglamento General de Protección de Datos (RGPD) (Considerando 26) aunque pudiera estar bajo el ámbito de aplicación de otras normas (p. ej. de seguridad nacional, salud pública, infraestructuras críticas, etc.) En este caso debe tenerse en cuenta que:

- **El tratamiento que generan los datos anonimizados sí es un tratamiento de datos personales, que puede considerarse compatible con el fin original del tratamiento de datos personales del que proceden los datos (Dictamen 05/2014 sobre técnicas de anonimización WP246 apartado 2.2.1. Legitimación del proceso de anonimización).**
- **El conjunto de datos anonimizados queda fuera del ámbito de aplicación del RGPD en la medida que es posible demostrar objetivamente que no existe capacidad material para asociar los datos anonimizados a una persona física determinada, directa o indirectamente, ya sea mediante el uso de otros conjuntos de datos, informaciones o medidas técnicas y materiales que pudieran existir a disposición de terceros.**

Es decir, los datos se considerarán anonimizados en la medida que no exista una probabilidad razonable que cualquier persona pueda identificar a la persona física en el conjunto de datos. Dicha evaluación ha de tener en cuenta los costes, el tiempo requerido para llevar a cabo

la reidentificación o los medios tecnológicos necesarios para conseguir la reversión de la anonimización, tanto los actuales como teniendo en cuenta los avances tecnológicos (Considerando 26).

El conjunto de datos seudonimizados, y la información adicional vinculada con dicho conjunto de datos, están bajo el ámbito de aplicación del RGPD, así como el tratamiento que los genera.

Los derechos y libertades de los interesados han de estar igualmente protegidas tanto en los tratamientos de anonimización como en los procesos de seudonimización. Teniendo en cuenta que sobre el conjunto de datos anonimizados no será preciso atender a los requisitos establecidos por el RGPD en cuanto a la limitación del tratamiento, la conservación de los datos, las comunicaciones y las transferencias internacionales, o las medidas para proteger la confidencialidad, se han de diseñar y validar los tratamientos de anonimización pensando en la protección de los derechos anteriormente señalados. Esto exige poder demostrar un nivel objetivo de calidad en el tratamiento de anonimización y aconseja determinar cómo evoluciona el riesgo de reidentificación a lo largo del tiempo. En cualquier caso, la reversión de la anonimización supone la plena aplicación del RGPD a los sujetos obligados que traten los datos personales.

"LOS DATOS ANÓNIMOS, NO PUEDEN ASOCIARSE CON UN INDIVIDUO EN PARTICULAR. UNA VEZ QUE LOS DATOS SON REALMENTE ANÓNIMOS Y LOS INDIVIDUOS DEJAN DE SER IDENTIFICABLES, DEJAN DE ESTAR INCLUIDOS EN EL ÁMBITO DE APLICACIÓN DEL RGPD"

LA SEUDONIMIZACIÓN NO ES LO MISMO QUE LA ANONIMIZACIÓN

La seudonimización no es lo mismo que la anonimizacion». El RGPD define «seudonimización» como «el tratamiento de datos personales de manera que no puedan atribuirse a un interesado sin utilizar información adicional, siempre que dicha información adicional figure por separado y esté sujeta a medidas técnicas y organizativas destinadas a garantizar que los datos personales no se atribuyan a una persona física identificada o identificable». Esto significa que el uso de «información adicional» puede suponer la identificación de los individuos; por ese motivo los datos personales seudonimizados también son datos personales.

Por el contrario, los datos anónimos, no pueden asociarse con un individuo en particular. Una vez que los datos son realmente anónimos y los individuos dejan de ser identificables, dejan de estar incluidos en el ámbito de aplicación del RGPD.

La seudonimización consiste en la sustitución de un atributo (normalmente un atributo único) por otro en un registro. Por consiguiente, sigue existiendo una alta probabilidad de identificar a la persona física de manera indirecta; en otras palabras, el uso exclusivo de la seudonimización no garantiza un conjunto de datos anónimo. No obstante, el presente dictamen examina este método debido a las numerosas ideas falsas y errores existentes sobre él. La seudonimización reduce la vinculabilidad de un conjunto de datos con la identidad del interesado; se trata, por tanto, de una medida de seguridad útil, pero no es un método de anonimización.

El resultado de la seudonimización puede ser independiente del valor inicial (tal sería el caso de un número aleatorio generado por el responsable del tratamiento o de un apellido escogido por el interesado) o bien derivarse de los valores originales de un atributo o conjunto de atributos, como por ejemplo en el caso de funciones hash o sistemas de cifrado.

"EL HECHO DE REDUCIR LA INFORMACIÓN EXISTENTE A LOS MÍNIMOS NECESARIOS PARA SATISFACER LOS OBJETIVOS QUE DEBE CUMPLIR LA INFORMACIÓN ANONIMIZADA IMPLICA DE FORMA DIRECTA UNA REDUCCIÓN DEL RIESGO DE REIDENTIFICACIÓN, PUES A MENOR CANTIDAD DE DATOS PERSONALES MENOR SERÁ EL RIESGO INHERENTE QUE RESULTE DEL TRATAMIENTO REALIZADO DURANTE EL PROCESO DE ANONIMIZACIÓN"

ASPECTOS A TENER EN CUENTA PARA ABORDAR LA ELIMINACIÓN DE LA IDENTIFICACIÓN EN LA ANONIMIZACIÓN

El objeto es reducir al mínimo necesario la cantidad de variables que permitan la identificación de las personas, restringiendo el acceso a la información confidencial al equipo de trabajo implicado en el proceso y optimizando el coste computacional de las operaciones con datos anonimizados. El hecho de reducir la información existente a los mínimos necesarios para satisfacer los objetivos que debe cumplir la información anonimizada implica de forma directa una reducción del riesgo de reidentificación, pues a menor cantidad de datos personales menor será el riesgo inherente que resulte del tratamiento realizado durante el proceso de anonimización: vulneración del deber de secreto, pérdida de información, brechas de seguridad, robo de claves, etc.

Algunos aspectos que pueden ser tenidos en cuenta por el responsable del tratamiento para abordar la eliminación o enmascarar las variables de identificación pueden ser los siguientes:

- Determinar la finalidad de los datos anonimizados: plazos de conservación, uso estadístico, científico, o cualquier uso posterior.
- Establecer las variables confidenciales necesarias para el tratamiento de los datos anonimizados e identificar las variables de confidencialidad que no vayan a ser necesarias en el tratamiento de los datos anonimizados. Como variable de confidencialidad se entenderá cualquier información existente sobre una persona, tanto si permite su identificación como si a priori su identificación no es posible.
- Eliminación de datos identificativos directos o indirectos no necesarios: nombres, fecha de nacimiento, teléfono, DNI, email, dirección postal, número de cuentas bancarias, matrículas de vehículos, identificador dispositivo móvil, número de serie, dirección IP, identificadores biométricos, fotografía o imagen, etc.
- Control segregado de usuarios con acceso a los datos personales y usuarios con acceso a los datos anonimizados: la información, anonimizada o no, puede estar estructurada y cada usuario tiene acceso a los datos que le son necesarios para realizar su trabajo, de forma que

"EL ALGORITMO DE HASH GENERA UNA HUELLA DIGITAL Y HACE IMPOSIBLE RECONSTRUIR EL DATO ORIGINAL PARTIENDO DE LA HUELLA Y POR OTRA PARTE CUALQUIER VARIACIÓN EN EL DATO ORIGINAL DARÁ LUGAR A UNA HUELLA DIGITAL DIFERENTE, LO QUE EXPRESADO EN TÉRMINOS COMPUTACIONALES PODRÍA DECIRSE QUE LA MODIFICACIÓN DE UN SOLO BIT EN LA INFORMACIÓN ORIGINAL ALMACENADA EN UN ORDENADOR DARÍA LUGAR A UNA CLAVE DISTINTA O UNA HUELLA DIGITAL DISTINTA. "

el resto de información o variables que pudieran permitir la reidentificación de los sujetos y no sean necesarias para el desempeño de las funciones de un trabajador no le sean accesibles.

- Utilización de rangos para enmascarar a las personas cuando existen microdatos concretos que permiten la identificación directa de personas o colectivos específicos. Por ejemplo, en el caso de colectivos de personas extremadamente reducidos se debe diluir la información de este pequeño grupo de personas en un colectivo de mayor rango numérico añadiendo, si es necesario, una referencia a un porcentaje en el que se ponga de manifiesto la existencia del colectivo menor como parte de un conjunto mayor. – Disponer de una política de uso de claves para ocultar la identificación de las personas será de gran utilidad en esta fase de la anonimización. La política de claves establecerá los niveles y el número de claves mínimo a tener en cuenta en función del objeto de la información anonimizada.

TECNICAS DE ANONIMIZACIÓN: CLAVES

ALGORITMOS DE HASH: es incuestionable la utilidad que tienen los algoritmos de cifrado cuando necesitamos anonimizar microdatos, resultando especialmente útiles los algoritmos de "hash". Un algoritmo de hash es un mecanismo que, aplicado a un dato concreto, genera una clave única o casi única que puede utilizarse para representar un dato. Por ejemplo, disponemos de un dato que queremos ocultar o anonimizar y para ello utilizamos un algoritmo de hash, como por ejemplo SHA1 o MD5. De la aplicación del algoritmo a un determinado dato obtenemos una clave o huella digital que puede utilizarse para reemplazar el dato real. El algoritmo de hash genera una huella digital y hace imposible reconstruir el dato original partiendo de la huella y por otra parte cualquier variación en el dato original dará lugar a una huella digital diferente, lo que expresado en términos computacionales podría decirse que la modificación de un solo bit en la información original almacenada en un ordenador daría lugar a una clave distinta o una huella digital distinta.

Sin embargo, un algoritmo de hash por sí solo no es suficiente para hacer irreversible la anonimización, ya que pequeñas cadenas de texto como, por ejemplo, los

microdatos correspondientes al código postal de una persona, un número de teléfono, etc., pueden ser fácilmente reidentificables con un programa informático que genere cifras consecutivas y sus correspondientes huellas digitales. Si lo que queremos es garantizar la anonimización de un microdato es preciso utilizar un mecanismo criptográfico que nos garantice el secreto de la huella digital que hemos generado. Una buena opción es el algoritmo HMAC basado en RFC20147 HMAC puede utilizarse en combinación con varios algoritmos de hash como, por ejemplo, con MD5 y sobre la huella digital o clave resultante del algoritmo de hash aplica un algoritmo criptográfico que genera una nueva huella digital o clave en función de una clave secreta.

ALGORITMOS DE CIFRADO: los algoritmos de cifrado con propiedades homomórficas abren nuevas posibilidades para el tratamiento de datos anonimizados. Un algoritmo de cifrado homomórfico permite realizar operaciones con datos cifrados de tal manera que el resultado de las operaciones es el mismo que si las operaciones se hubieran realizado con los datos sin cifrar. Los resultados de las operaciones con datos cifrados dan por resultado valores igualmente cifrados que pueden ser descifrados posteriormente por el usuario que disponga de la clave para descifrar. El esquema de cifrado homomórfico abre la posibilidad del tratamiento de datos personales anonimizados garantizando la privacidad del tratamiento y que los resultados de los tratamientos van a ser accesibles únicamente al poseedor de la clave de descifrado.

La implantación de esquemas de cifrado homomórfico puede aportar un alto grado de confidencialidad a los tratamientos de datos en cloud, tratamientos de datos obtenidos de los wearables, sistemas de telemedicina, etc.

SELLO DE TIEMPO: también hay que 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 la anonimización ha sido realizada, o incluso algoritmos de firma electrónica que permiten garantizar la identidad electrónica de quien ha realizado la anonimización.

"LOS CRITERIOS PARA LA ANONIMIZACIÓN POR CAPAS PUEDEN FIJARSE EN RESPUESTA A REQUISITOS DE LA INFORMACIÓN, DEL TRATAMIENTO DE LOS DATOS ANONIMIZADOS O DE LA PROPIA POLÍTICA DE ANONIMIZACIÓN DEL RESPONSABLE DEL TRATAMIENTO DE LOS DATOS ANONIMIZADOS, ATENDIENDO A DIFERENTES CRITERIOS SE PODRÁN ABORDAR DISTINTAS TÉCNICAS."

CAPAS DE ANONIMIZACIÓN: junto con estos procesos de enmascaramiento y anonimización podemos utilizar lo que podría denominarse la anonimización por capas. Por ejemplo, el responsable del tratamiento ha anonimizado todos los datos que puedan servir para reidentificar a las personas y remite la información a su legítimo destinatario quien, a fin de evitar que pudiera producirse la reidentificación, decide realizar una segunda anonimización de los datos ya anonimizados. De esta forma, el destinatario de la información anonimizada asegura que sus procesos utilizan sus propios recursos de anonimización, evitando que en caso de fragilidad de los procesos de anonimización del responsable del tratamiento la identidad de las personas pudiera verse afectada.

En definitiva, el destinatario de la información asegura la privacidad de las personas en el tratamiento de datos anonimizados con sus propias garantías de calidad. Este proceso de reanonimización puede utilizarse de manera interdepartamental de forma que, a medida que los datos vayan pasando de un departamento a otro, se realicen diferentes procesos de anonimización haciendo que para una reidentificación real sea necesaria la concurrencia de todos de los actores involucrados en las distintas capas de anonimización.

Puede decirse, por lo tanto, que el proceso de anonimización puede ser monocapa o multicapa. Hablamos de un proceso de anonimización monocapa cuando la anonimización de las variables se realiza una única vez y se da por finalizado el proceso pero, en ocasiones, la reanonimización de variables o anonimización multicapa puede proporcionar garantías adicionales para evitar la reidentificación de las personas.

Los criterios para la anonimización por capas pueden fijarse en respuesta a requisitos de la información, del tratamiento de los datos anonimizados o de la propia política de anonimización del responsable del tratamiento de los datos anonimizados, atendiendo a diferentes criterios se podrán abordar distintas técnicas.

GARANTÍAS DE LAS TECNICAS DE ANONIMIZACIÓN

En relación a las técnicas de anonimización, el Grupo de Trabajo del artículo 29 emitió su Dictamen 05/2014 en el que se realiza un análisis técnico acerca de la robustez, debilidad y garantías de las técnicas de anonimización. En dicho Dictamen se muestran algunos de los límites, riesgos y errores que pueden tener lugar como resultado de las técnicas de anonimización utilizadas.

Es preciso tener en cuenta que la anonimización de la información siempre generará, independientemente de las buenas prácticas empleadas, cierto grado de distorsión entre la información anonimizada y la información no anonimizada, y que a veces el proceso de anonimización no puede asegurar la imposibilidad de reidentificación de las personas en términos absolutos, por lo cual se deben de tener en cuenta las garantías jurídicas necesarias para preservar los derechos de los interesados.

En este sentido algunos de los aspectos que deben ser tenidos en cuenta, a juicio de la AEPD, son:

- **Acuerdos de confidencialidad** que impliquen a los siguientes actores:
- Responsable del tratamiento.
- Responsable del proceso de anonimización.
- Responsable del tratamiento de datos anonimizados.
- Personal con acceso a la información anonimizada.
- Obtener el **compromiso del destinatario** de la información para mantener la anonimización y la obligación de informar al responsable del tratamiento ante cualquier sospecha de reidentificación.
- Realización de **auditorías de uso de la información anonimizada** por parte del responsable del tratamiento al responsable del tratamiento de los datos anonimizados.
- Las garantías estarán **incluidas en el contrato suscrito entre el responsable del tratamiento y el destinatario** de la información anonimizada.

MATERIAL COMPLEMENTARIO

- Dictamen 05/2014 del GRUPO DE TRABAJO SOBRE PROTECCIÓN DE DATOS DEL ARTÍCULO 29. Consulta este enlace.
- "Orientaciones y garantías en los procedimientos de anonimizacion de datos personales" GUIA (AEPD). Consulta este enlace.
- "10 Malentendidos relacionados con la anonimizacion" (AEPD). Consulta este enlace.
- Blog de la AEPD "Anonimización y seudonimización". Consulta este enlace.
- Blog de la AEPD "Anonimización y seudonimización (II): la privacidad diferencial". Consulta este enlace.
- Resolución Procedimiento Nº: PS/00052/2020 en relación con un aplicativo que muestra datos que deberían ser anonimizados (AEPD). Consulta este enlace.

NOTICIAS

- Entra en vigor el Real Decreto-ley 24/2021, de 2 de noviembre, que transpone directivas de la Unión Europea en determinadas materias como la relativa a los datos abiertos y la reutilización de la información del sector público. El Libro tercero del incorpora al ordenamiento jurídico español las novedades de la Directiva (UE) 2019/1024 del Parlamento Europeo y del Consejo, de 20 de junio de 2019, relativa a los datos abiertos y la reutilización de la información del sector público. Consulta la norma en este enlace.
- La AEPD sanciona a la Consejería de Educación del Principado de Asturias por no utilizar protocolo seguro en su web y por incorporar un apartado de "Trabaja con nosotros" sin dar cumplimiento a la Ley de Servicios de la Sociedad de la Información. En el procedimiento instruido por la AEPD, se sanciona a la citada Consejería por entender que la página no es segura, pues utiliza el protocolo http://; así como por considerar que no cumple con la Ley de Servicios de la Sociedad de la Información, al considerarla "prestador de servicios de la sociedad de la información", pues gestionaba la bolsa de trabajo del Instituto, Consulta la Resolución en este enlace.
- La AEPD sanciona a la Consejería de Educación y Juventud de la Comunidad de Madrid por la grabación telemática de las sesiones de evaluación de un IES. Los docentes que no pudieron asistir a la sesión presencial participaron de forma telemáticamente. Tales sesiones fueron grabadas por decisión de la dirección del centro educativo. En el presente caso, la AEPD considera acreditado el incumplimiento del principio de trasparencia (por la falta de información a los profesores participantes), así como el principio de licitud del tratamiento. Consulta la Resolución en este enlace.