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 N.º 38| Agosto 2023

ERRORES COMUNES EN MATERIA DE ANONIMIZACIÓN

ÍNDICE

ERRORES COMUNES EN MATERIA DE ANONIMIZACIÓN

Os invitamos a trasladarnos aquellas temáticas que resulten de vuestro interés para los próximos cuadernos. 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 suscribirse a nuestro Cuaderno accede al siguiente enlace.

"Los datos anónimos constituyen: «aquella información que no hace referencia a personas naturales identificadas o identificables o a datos personales que se anonimizan de tal forma que dejan de ser identificables»".

ÍNTRODUCCIÓN

La **información anónima** es un conjunto de datos que no guarda relación con una persona física identificada o identificable (Considerando 26 del Reglamento (UE) 2016/679 General de Protección de Datos, en adelante, 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 del RGPD).

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.

La anonimización, sus técnicas, procesos y resultados, son cuestiones complejas que han generado ciertas dudas en tanto su aplicación práctica.

Por ello, a través del presente documento se analizan y resuelven las principales confusiones que ha generado la técnica de la anonimización, explicando la realidad y ofreciendo referencias para una lectura pormenorizada.

Estos malentendidos han sido resueltos por la Agencia Española de Protección de Datos (en adelante, AEPD) en su documento llamado: 10 MALENTENDIDOS RELACIONADOS CON LA ANONIMIZACIÓN.

"REALIDAD: «La seudonimización no es lo mismo que la anonimización»".

ERROR 1: «La seudonimización es lo mismo que la anonimización»

La realidad ante esta cuestión es que la seudonimización **no es lo mismo** que la anonimización.

El RGPD define «*seudonimización*» en su artículo 4.5) 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.

ERROR 2: «El cifrado es anonimización»

La realidad se encuentra en que **el cifrado no constituye una técnica de anonimización**, pero puede ser una buena herramienta de seudonimización.

El proceso de cifrado utiliza claves secretas para transformar la información de tal forma que se reduzca el riesgo de uso indebido y, al mismo tiempo, se mantenga la confidencialidad durante un periodo de tiempo determinado. Dado que la información original debe ser accesible, las transformaciones aplicadas por los algoritmos de cifrado están diseñadas para ser reversibles, lo que se conoce como descifrado.

Las claves privadas que se utilizan para el descifrado son la «*información adicional*», mencionada anteriormente (véase Error 1), lo que puede hacer que los datos sean legibles y, en última instancia, que la identificación sea posible.

En teoría, podría considerarse que la eliminación de la clave de cifrado de los datos cifrados los convertiría en anónimos, pero no es así. No se puede dar por hecho que los datos cifrados no puedan descifrarse porque se diga que la clave de descifrado se ha «*borrado*» o es «*desconocida*». Hay muchos factores que afectan a la confidencialidad de los datos cifrados, en particular a largo plazo. Entre estos factores se encuentran la solidez del algoritmo de cifrado y de la clave, las fugas de información, los problemas de implantación, la cantidad de datos cifrados o los avances tecnológicos (por ejemplo, la computación cuántica).

"REALIDAD: El cifrado no constituye una técnica de anonimización, pero puede ser una buena herramienta de seudonimización".

ERROR 3: «Los datos siempre pueden anonimizarse»

La realidad es que **no siempre es posible reducir el riesgo de reidentificación** por debajo de un umbral definido de forma previa y mantener, al mismo tiempo, la utilidad de un conjunto de datos para un tratamiento específico.

La anonimización es un proceso que trata de encontrar el equilibrio adecuado entre la reducción del riesgo de reidentificación y el mantenimiento de la utilidad de un conjunto de datos para los fines previstos. Sin embargo, en función del contexto o la naturaleza de los datos, los riesgos de reidentificación podrían no mitigarse lo suficiente. Esta situación puede darse cuando el número total de posibles individuos («*universo de sujetos*») es demasiado reducido (por ejemplo, un conjunto de datos anónimos que contenga sólo los 705 miembros del Parlamento Europeo), cuando las categorías de datos son tan diferentes entre los individuos que es posible individualizarlos (por ejemplo, la huella digital del dispositivo de los sistemas que accedieron a un determinado sitio web) o cuando el caso de los conjuntos de datos incluye un elevado número de atributos demográficos o datos de localización.

"REALIDAD: No siempre es posible reducir el riesgo de reidentificación por debajo de un umbral definido de forma previa y mantener, al mismo tiempo, la utilidad de un conjunto de datos para un tratamiento específico".

ERROR 4: «La anonimización es permanente»

La realidad es que **existe un riesgo de que ciertos procesos de anonimización puedan revertirse en el futuro.** Las circunstancias pueden cambiar a lo largo del tiempo y los nuevos avances técnicos y la disponibilidad de información adicional pueden poner en peligro los procesos de anonimización previos.

Los recursos informáticos y las nuevas tecnologías (o nuevos usos de tecnologías ya existentes) disponibles para un atacante que pudiera intentar reidentificar un conjunto de datos anónimos van cambiando a lo largo del tiempo. Hoy en día, la computación en la nube proporciona una capacidad de computación asequible a niveles y precios que eran impensables hace años. En el futuro, los ordenadores cuánticos también podrían alterar lo que en la actualidad se consideran «*medios aceptables*».

Además, la divulgación de datos adicionales a lo largo de los años (por ejemplo, en una filtración de datos personales) puede permitir que los datos que anteriormente eran anónimos se atribuyan a personas identificadas. La divulgación de registros de muchas décadas de antigüedad que contengan datos muy sensibles (por ejemplo, antecedentes penales) podría continuar teniendo un efecto bastante perjudicial para un individuo o sus familiares.

"REALIDAD: Existe un riesgo de que ciertos procesos de anonimización puedan revertirse en el futuro. Las circunstancias pueden cambiar a lo largo del tiempo y los nuevos avances técnicos y la disponibilidad de información adicional pueden poner en peligro los procesos de anonimización previos."

ERROR 5: «La anonimización siempre reduce la probabilidad de reidentificación de un conjunto de datos a cero»

La realidad se encuentra en que **el proceso de anonimización y la forma en que se aplique** tendrán una influencia directa en la probibilidad de **riesgos de reidentificación.**

Un proceso de anonimización sólido tiene como objetivo la reducción del riesgo de reidentificación por debajo de un determinado umbral. Dicho umbral dependerá de varios factores, como los controles de mitigación existentes (ninguno en el contexto de la divulgación pública), la repercusión en la privacidad de los individuos en caso de reidentificación, los motivos y la capacidad de un atacante para reidentificar los datos.

Aunque una anonimización del 100% es el objetivo más deseable desde el punto de vista de la protección de los datos personales, en algunos casos no es posible y debe contemplarse un riesgo residual de reidentificación.

"REALIDAD: El proceso de anonimización y la forma en que se aplique tendrán una influencia directa en la probabilidad de riesgos de reidentificación".

"REALIDAD: El grado de anonimización puede analizarse y medirse."

ERROR 6: «La anonimización es un concepto binario que no puede medirse»

La realidad es que **el grado de anonimización puede analizarse y medirse.**

La expresión «*datos anónimos*» no puede entenderse como si los conjuntos de datos pudieran etiquetarse como anónimos o no. Existe una probabilidad de que los registros de cualquier conjunto de datos se reidentifiquen en función de la posibilidad de individualizarlos. Cualquier proceso sólido de anonimización evaluará el riesgo de reidentificación, que debe gestionarse y controlarse a lo largo del tiempo.

Excepto en casos específicos en los que los datos estén muy generalizados (por ejemplo, un conjunto de datos que cuente el número de visitantes de un sitio web por país en un año), el riesgo de reidentificación nunca puede considerarse nulo.

ERROR 7: «La anonimización puede automatizarse totalmente»

La realidad se encuentra en que es posible utilizar herramientas automáticas durante el proceso de anonimización, pero, dada la importancia del contexto en la evaluación de dicho proceso, **la intervención del experto humano es necesaria.**

Al contrario, requiere un análisis del conjunto de datos original, sus fines previstos, las técnicas que deben aplicarse y el riesgo de reidentificación de los datos resultantes.

Pese a que la identificación y eliminación de los identificadores directos (también conocida como «enmascaramiento») constituye una parte importante del proceso de anonimización, debe ir siempre acompañada de un análisis cauteloso que busque otras fuentes de identificación (indirecta, en general, a través de cuasiidentificadores). Mientras que encontrar los identificadores directos es algo trivial, los identificadores indirectos, en cambio, no siempre son obvios, y el hecho de no detectarlos puede dar lugar a la reversión del proceso (es decir, la reidentificación), lo que tiene consecuencias para la privacidad de los individuos.

La automatización podría ser clave en algunos pasos del proceso de anonimización, como la eliminación de identificadores directos o la aplicación coherente de un procedimiento de generalización sobre una variable. Por el contrario, parece poco probable que un proceso totalmente automatizado pueda identificar cuasiidentificadores en diferentes contextos o decidir cómo maximizar la utilidad de los datos aplicando técnicas específicas a variables concretas.

"REALIDAD: Es posible utilizar herramientas automáticas durante el proceso de anonimización, pero, dada la importancia del contexto en la evaluación de dicho proceso, la intervención del experto humano es necesaria."

"REALIDAD: Un proceso de anonimización adecuado mantiene la funcionalidad de los datos para un fin determinado".

ERROR 8: «La anonimización inutiliza los datos»

La realidad es que un proceso de anonimización adecuado **mantiene la funcionalidad de los datos para un fin determinado.**

El objetivo de la anonimización es evitar que se identifique a los individuos de un conjunto de datos. Las técnicas de anonimización siempre restringirán las formas en que se puede utilizar el conjunto de datos resultante. Por ejemplo, agrupar las fechas de nacimiento en intervalos de un año reducirá el riesgo de reidentificación y, al mismo tiempo, la utilidad del conjunto de datos en algunos casos. Esto no significa que los datos anónimos sean inútiles, sino que su utilidad dependerá de la finalidad y del riesgo de reidentificación que se acepte.

Por otra parte, los datos personales no pueden almacenarse de forma permanente más de lo que estipule su finalidad original, a la espera de una oportunidad en la que puedan resultar útiles para otros fines. La solución para algunos responsables del tratamiento podría ser la anonimización, en la que los datos personales pueden independizarse y desecharse del conjunto de datos, mientras que el conjunto de datos restante continúa conservando un significado útil.

El principio de «*minimización de los datos*» exige que el responsable del tratamiento determine si es necesario tratar los datos personales para cumplir un objetivo concreto, o si ese objetivo puede alcanzarse también con datos anónimos.

En algunos casos, esto puede conducir a la conclusión de que la anonimización de los datos no se ajusta a la finalidad prevista. En estos casos, el responsable del tratamiento tendrá que decidir entre tratar los datos personales (y utilizar, por ejemplo, la seudonimización) y aplicar el RGPD, o no tratar los datos de ninguna forma.

ERROR 9: «Seguir un proceso de anonimización que otros utilizaron con éxito hará que nuestra organización obtenga resultados equivalentes»

La realidad se encuentra en que **los procesos de anonimización deben adaptarse al caso concreto,** esto es, a la naturaleza, el alcance, el contexto y los fines del tratamiento, así como a los riesgos de diversa probabilidad y gravedad para los derechos y libertades de las personas físicas.

La anonimización no puede aplicarse como si se siguiera una receta, porque el contexto (naturaleza, alcance, contexto y fines del tratamiento de los datos) probablemente difiera de una circunstancia a otra y de una organización a otra. Un proceso de anonimización puede tener un riesgo de reidentificación por debajo de un determinado umbral cuando los datos sólo se ponen a disposición de un número limitado de destinatarios, mientras que el riesgo de reidentificación no podrá alcanzar ese umbral cuando los datos se pongan a disposición del público en general.

Puede haber diferentes conjuntos de datos disponibles en diferentes contextos. Estos podrían cruzarse con los datos anónimos, lo que afectaría al riesgo de reidentificación. Por ejemplo, en Suecia, la información relativa a los datos personales de los contribuyentes está disponibles de forma pública, mientras que en España no lo están. Por tanto, aunque los conjuntos de datos que incluyen información de ciudadanos españoles y suecos se anonimizaran siguiendo el mismo procedimiento, los riesgos de reidentificación podrían ser diferentes.

"REALIDAD: Los procesos de anonimización deben adaptarse a la naturaleza, el alcance, el contexto y los fines del tratamiento, así como a los riesgos de diversa probabilidad y gravedad para los derechos y libertades de las personas físicas".

ERROR 10: «No existe un riesgo ni interés alguno en saber a quién se atribuyen estos datos»

La realidad es que los datos personales tienen un **valor** en sí mismos, para los propios individuos y para terceros. **La reidentificación de un individuo podría tener una repercusión grave en lo relativo a sus derechos y libertades.**

Los ataques contra la anonimización pueden materializarse en forma de intentos deliberados de reidentificación, intentos involuntarios de reidentificación, brechas de seguridad o divulgación de datos al público. La probabilidad de que alguien intente reidentificar a un individuo solo se refiere al primer tipo. No se puede descartar la posibilidad de que alguien reidentifique al menos a una persona en un conjunto de datos, ya sea por curiosidad, por casualidad o por un interés real (por ejemplo, investigación científica, periodismo o actividad delictiva).

Puede ser difícil evaluar con precisión el impacto de la reidentificación en la vida privada de una persona, porque siempre dependerá del contexto y de la información que se correlacione. Por ejemplo, la reidentificación de un interesado en el contexto aparentemente inofensivo de sus preferencias cinematográficas podría llevar a inferir sobre las inclinaciones políticas o la orientación sexual de esa persona.

Sin embargo, estos datos especialmente sensibles gozan de una protección especial en virtud del RGPD.

"REALIDAD: Los datos personales tienen un valor en sí mismos, para los propios individuos y para terceros. La reidentificación de un individuo podría tener una repercusión grave en lo relativo a sus derechos y libertades".

MATERIAL COMPLEMENTARIO

- 10 Malentendidos relacionados con la anonimización publicado por la Agencia Española de Protección de Datos (AEPD) en este enlace.
- Guía básica de anonimización Elaborada por Autoridad Nacional de Protección de Datos de Singapur (PDPC - Personal Data Protection Commission Singapore) y traducida por la Agencia Española de Protección de datos (AEPD) en este enlace.
- Orientaciones y garantías en los procedimientos de anonimización de datos personales de la Agencia Española de Protección de datos (AEPD) en este enlace.

NOTICIAS

Las autoridades en protección de datos publican varias noticias de interés:

1. **La APDCAT pone en marcha la red de delegados y delegadas de protección de datos en Cataluña.**

   La Autoridad Catalana de Protección de Datos ha presentado *'DPD en xarxa'*, la comunidad de aprendizaje y colaboración de delegados y delegadas datos de Cataluña, impulsada para favorecer la formación, la comunicación, el intercambio de experiencia y el trabajo en grupo entre las personas que velan por que se cumpla la normativa de protección de datos en las organizaciones del sector público catalán.

   Consulta la publicación en este enlace.

2. **El proyecto de nueva Ley Vasca de Protección de Datos personales.**

   El **Proyecto de Ley de Protección de Datos Personales y de la Autoridad Vasca de Protección de Datos**, en proceso de tramitación en el Parlamento Vasco, ha sido aprobado con el objetivo de regular el control y la supervisión del tratamiento de los datos personales de los que sean responsables los sujetos incluidos en su ámbito de aplicación.

   El texto presenta como principal novedad la regulación de la "**Autoridad Vasca de Protección de Datos"**, que sustituirá a la actual "Agencia Vasca de Protección de Datos".

   Consulta la publicación en este enlace.