Orientaciones y garantías en los procedimientos de ANONIMIZACIÓN de datos personales

AGENCIA ESPAÑOLA DE PROTECCIÓN DE DATOS

Orientaciones y garantías en los procedimientos de ANONIMIZACIÓN de datos personales

AGENCIA ESPAÑOLA DE PROTECCIÓN DE DATOS

índice

El diseño de la cadena de confidencialidad supone el estudio y análisis específicos para cada bloque de datos personales, que deberán adecuarse para cada situación específica y cada bloque de datos que se pretende anonimizar. En este documento se facilitan algunas orientaciones que pueden ser tenidas en cuenta por los responsables del tratamiento de los datos personales, incentivando a los responsables a la aplicación de procedimientos de anonimización.

Los procesos y técnicas descritos en estas orientaciones pueden resultar útiles tanto para los procesos de anonimización propiamente dichos como para reforzar la seguridad de la información, aportando mayores garantías en los tratamientos y en la conservación de datos personales.

En este sentido, las medidas que se plantean pueden aplicarse tanto a datos de personas físicas, como a los datos de personas jurídicas como administraciones, empresas, centros educativos, asociaciones, etc. de los que se recoja y compile información y que por diversos motivos no deseen ser identificadas o hayan puesto el anonimato como condición para ceder esos datos para su publicación.

1. ANONIMIZACIÓN

La finalidad del proceso de anonimización es eliminar o reducir al mínimo los riesgos de reidentificación de los datos anonimizados manteniendo la veracidad de los resultados del tratamiento de los mismos, es decir, además de evitar la identificación de las personas, los datos anonimizados deben garantizar que cualquier operación o tratamiento que pueda ser realizado con posterioridad a la anonimización no conlleva una distorsión de los datos reales. Un análisis masivo de los datos o macrodatos que puedan derivar de los datos anonimizados no debería diferir del análisis que pudiera obtenerse si hubiera sido realizado con datos no anonimizados.

En el proceso de anonimización se deberá producir la ruptura de la cadena de identificación de las personas. Esta cadena se compone de microdatos o datos de identificación directa y de datos de identificación indirecta. Los microdatos permiten la identificación directa de las personas y los datos de identificación indirecta[1] son datos cruzados de la misma o de diferentes fuentes que pueden permitir la reidentificación de las personas, como la información de otras bases de datos del mismo u otro responsable, de las redes sociales, buscadores, blogs, etc.

En el diseño del proceso de anonimización será necesario prever las consecuencias de una eventual reidentificación de las personas que pudiera generar un perjuicio o merma de sus derechos. Igualmente 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.

[1] Identificación indirecta: la que puede tener lugar como consecuencia de información de una o varias fuentes que por sí misma o en combinación de otros factores puede permitir la reidentificación de las personas cuando sus datos hubieran sido anonimizados. Por ejemplo, la combinación de sexo, edad, lugar de nacimiento y padecimiento de una determinada enfermedad pueden permitir la identificación indirecta de una persona concreta.

2. PRINCIPIOS DE LA ANONIMIZACIÓN

A continuación se relacionan algunos principios a tener en cuenta en un proceso de anonimización.

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 serán tenidos en cuenta desde las etapas iniciales del diseño del sistema de información o del producto utilizado para el proceso de anonimización y durante todo el ciclo de vida de dicho producto o sistema de información. El concepto de privacidad desde el diseño en los procesos de anonimización puede resumirse en la aplicación de los siguientes principios:

1. **Principio proactivo.** La protección de la privacidad es el primer objetivo de la anonimización y su gestión debe realizarse de forma proactiva y no reactiva. Desde el inicio conceptual del diseño del sistema de información o producto a utilizar en el proceso de anonimización se tomarán las medidas necesarias para garantizar la privacidad de las personas. La privacidad no puede garantizarse a posteriori como el resultado de la reparación de brechas existentes en el proceso de anonimización o perjuicios ocasionados a los interesados, por lo que es necesario asegurar la inexistencia de posibles cadenas de reidentificación de los interesados en los datos anonimizados.

   Una medida importante en el estado inicial del concepto de un sistema de información o producto empleado en los procesos de anonimización es realizar una clasificación inicial de los datos y disponer de una escala o gradiente de sensibilidad de la información. Esta clasificación puede ser cualitativa o cuantitativa y servirá de referencia dentro de una organización.

   Por ejemplo, puede desarrollarse un esquema de clasificación consistente en un esquema basado al menos en tres niveles de identificación de personas (microdatos, datos de identificación indirecta y datos sensibles), donde se asigne un valor cuantitativo a cada una de las variables de identificación. La escala será conocida por todo el personal implicado en el proceso de anonimización y será clave fundamental a tener en cuenta en el análisis de riesgos o Evaluación de Impacto en la Protección de los Datos Personales (EIPD[2]) del proceso de anonimización.

2. **Principio de privacidad por defecto.** El primer requisito conceptual en el diseño de un sistema de información será garantizar la confidencialidad de los interesados. Por lo tanto, conviene que desde el inicio se salvaguarde la privacidad teniendo en cuenta la granularidad o grado de detalle final que deben tener los datos anonimizados. En este sentido, mantener una escala cualitativa o cuantitativa como la referida en el punto anterior es una herramienta de indudable utilidad para la eliminación de las variables atendiendo a criterios de granularidad preestablecidos.

[2] EIPD: Para más información sobre las evaluaciones de impacto en la privacidad puede consultarse la Guía para una Evaluación de Impacto en la Protección de Datos Personales de esta Agencia.

Este equipo sería el encargado de definir el proceso y los agentes implicados así como las funciones de cada uno de ellos, evitando que exista alguna tarea que, siendo necesaria para garantizar la confidencialidad, no hubiera sido asignada a un responsable concreto.

Este posible esquema de segregación de roles y tareas plantea, en definitiva, un posible inventario funcional y orgánico que podría servir de orientación a la hora de planificar los procesos de anonimización.

3.2. INDEPENDENCIA DE FUNCIONES

Como ya se ha indicado, sería recomendable que la definición del equipo de trabajo y de los diferentes actores implicados se realizara atendiendo al principio de independencia profesional, de forma que cada uno de ellos obrará en el ámbito de las funciones que le hayan sido asignadas y siempre que fuera posible se tratara de evitar que la misma persona pudiera estar implicada en varios perfiles o funciones diferentes.

Uno de los objetivos del principio de independencia de funciones es el de evitar que un error que se produzca en un determinado nivel sea supervisado y aprobado en otro nivel distinto por la misma persona. No obstante, las condiciones y los recursos con los que se pueda abordar un proceso de anonimización puede que no permitan una segregación exhaustiva de funciones y sea necesario que algunos de los integrantes del equipo de trabajo tengan que realizar varios roles diferentes.

Es fundamental que exista un documento de definición del equipo de trabajo en el que se acote de forma expresa cada una de las funciones de las personas implicadas. Este documento deberá ser aprobado por el responsable de la información y actualizado siempre que sea necesario por el responsable de la anonimización. En los casos que no sea posible la segregación y la independencia de funciones se hará constar en el documento los motivos o situaciones que condicionen la independencia o la segregación de funciones. El objetivo de este documento es garantizar que cada tarea encaminada a la anonimización o disociación definitiva de los datos personales tenga asociado un responsable.

3.3. EVALUACIÓN DE RIESGOS DE REIDENTIFICACIÓN

Los procesos de anonimización de datos tienen un impacto directo en los recursos de una organización. Este esfuerzo asociado a los procesos de anonimización (económico, tecnológico, humano, etc.) deberá ser adecuado a los objetivos y requerimientos del proceso de anonimización. El esfuerzo requerido dependerá de las medidas que se precise implantar para garantizar la privacidad de los datos personales, adecuando las medidas necesarias a la finalidad de los datos anonimizados y a los posibles riesgos de reidentificación de las personas.

Cuando se utilizan de forma exclusiva mecanismos de anonimización sin realizar una EIPD previa

uno de ellos. Algunos de los activos que debemos de tener en cuenta en la elaboración del análisis de riesgos son:

a. Datos personales a anonimizar.
b. Activos de información anonimizada y variables de identificación asociadas.
c. Procesos y subprocesos de anonimización.
d. Sistemas de información implicados: hardware utilizado, limitación del software de anonimización con relación a los activos de información que sea preciso anonimizar.
e. Análisis de dependencias de activos implicados en el proceso de anonimización.
f. Categorización de los activos: es posible establecer una categorización en función de la criticidad de cada activo, teniendo en cuenta aspectos como, por ejemplo, grado de sensibilidad de la información.

2. CONSTITUCIÓN DEL EQUIPO DE TRABAJO: el grado de especialización en análisis de riesgos y en protección de datos son dos factores a considerar en el momento que se constituye el equipo de trabajo. Será necesario tener en cuenta la participación de los representantes del equipo de trabajo, que en última instancia tendrán que aceptar la propuesta del riesgo residual aceptable.

3. IDENTIFICACIÓN DE RIESGOS: El primer aspecto a tener en cuenta en el análisis de riesgos es la catalogación inicial de riesgos atendiendo a tres categorías:

a. Riesgos de reidentificación existentes conocidos
b. Riesgos potenciales de reidentificación
c. Riesgos no conocidos

A cada riesgo identificado en las mencionadas categorías se le asignará un valor determinado de una escala cuantitativa o cualitativa en función de la probabilidad de ocurrencia. El conjunto de todos los riesgos dará lugar al catálogo de riesgos de un bloque de información que se pretenda anonimizar.

Algunos de los riesgos que pueden ser tenidos en cuenta son: los riesgos de reidentificación por correlación con otros conjuntos de datos, los riesgos de vulneración del deber de secreto por acceso indebido a la información sin anonimizar, riesgos de revelación de claves de anonimización de la información, la existencia de un atacante o adversario potencial o lo que puede considerarse como el rol del "perseguidor", o la existencia de un sujeto que conoce la identidad de una persona en un bloque de información y que pretende obtener mayor información, etc.

4. VALORACIÓN DE LOS RIESGOS EXISTENTES: atendiendo al conjunto de activos y al catálogo de riesgos existentes se realizará una categorización de cada uno de los riesgos de reidentificación que hubieran sido detectados. Esta categorización se tendrá presente en las fases del proceso de anonimización y especialmente cuando se realice la eliminación de variables.

recomendaciones a implantar en los procesos de anonimización y en la explotación de la información anonimizada.

11. REVISIÓN DE RIESGOS: el análisis de riesgos se debe realizar de forma periódica a lo largo de todo el ciclo de vida de la información y siempre que se produzcan cambios en los procesos de anonimización o en el tratamiento de la información anonimizada. Las revisiones periódicas tienen por finalidad verificar que el estado real de riesgos coincide con el riesgo asumible de reidentificación, verificando la eficacia de las medidas previstas para paliar el posible impacto que pudiera tener la reidentificación de las personas. Por su parte, el responsable del tratamiento de los datos anonimizados deberá tener en cuenta el catálogo de riesgos inicial elaborado por el responsable del tratamiento para continuar elaborando su propio análisis a lo largo del ciclo de vida de la información anonimizada.

3.4. DEFINICIÓN DE OBJETIVOS Y FINALIDAD DE LA INFORMACIÓN ANONIMIZADA

Además de garantizar la privacidad de los interesados, el responsable del tratamiento determinará los objetivos que deberá cumplir la información anonimizada en función de los legítimos intereses de su destinatario. El diseño del proceso de anonimización estará condicionado por el objetivo final de la información anonimizada dando lugar a información de uso restringido o a datos abiertos.
Cuando la información anonimizada tiene por finalidad convertirse en información de uso restringido, la privacidad de los datos personales será reforzada mediante acuerdos de confidencialidad que formarán parte del conjunto de las garantías jurídicas del proceso de anonimización. En el caso de información anonimizada de uso restringido, el responsable del tratamiento podrá valorar la elaboración de posibles cláusulas contractuales, códigos de conducta y mecanismos de certificación que incluyan el compromiso por parte del destinatario de no realizar ningún intento para la reidentificación de las personas y que garanticen la privacidad de la información incluso cuando se produzcan brechas de reidentificación.

3.5. VIABILIDAD DEL PROCESO

Cuando se pretenda anonimizar datos especialmente protegidos a los que se refiere el artículo 9 del RGPD, se podría tener en cuenta la existencia de un equipo para el estudio de la viabilidad del proceso de anonimización. La labor de este equipo tendrá especial relevancia y su tarea principal sería la realización de un informe de viabilidad que reflejará detalladamente los motivos y condiciones específicas para la anonimización de los datos especialmente protegidos. En dicho informe podrían incluirse entre otros, por ejemplo, fundamentos o vinculaciones éticas del proceso de anonimización.

Los expertos en seguridad ejercerán como órgano consultivo con el fin de aportar viabilidad técnica a los fundamentos que habiliten la utilización de la información anonimizada y el proceso de anonimización. De manera conjunta darán conformidad al umbral de riesgos aceptable que resulte de la EIPD y, en caso de no hacerlo, deberán emitir el correspondiente dictamen motivado.

3.6. PREANONIMIZACIÓN: DEFINICIÓN DE VARIABLES DE IDENTIFICACIÓN

La preanonimización de los microdatos es la parte inicial del proceso de anonimización, en el que se determinarán las posibles variables de identificación (directas e indirectas) a tener en cuenta en el diseño de las herramientas de anonimización. Durante el proceso de preanonimización se tendrá en cuenta:

- La determinación de variables: datos personales, identificadores directos e indirectos[4], datos especialmente protegidos y otros datos con carácter confidencial.
- La clasificación y sensibilidad de las variables por categorías: de identificación directa, de identificación geográfica, de carácter especialmente protegido, numéricas, temporales, metadatos, etc.
- Variables de identificación que no puedan ser anonimizadas y que sea preciso eliminar del proceso de anonimización.
- Variables anonimizadas que sean imprescindibles para la finalidad a la que se van a destinar los datos anonimizados.

Una vez que se ha realizado una categorización de las variables se establecen los criterios de protección necesarios para garantizar la privacidad de las personas, tratando de minimizar la cantidad de información personal que vaya a ser utilizada durante el proceso de anonimización.
No podrá abordarse el proceso de anonimización de variables sin antes definir las posibles variables de identificación que vayan a ser realmente necesarias para la finalidad a la que será destinada la información anonimizada. En este proceso existen variables o microdatos que son elementos de identificación tangible, pero existen otras variables de identificación indirecta que permiten la identificación de las personas de forma menos tangible. Este hecho redunda en la dificultad del proceso de identificación de variables.

Partiendo de un conjunto de datos que han sido recopilados de los interesados siguiendo los principios de calidad de datos que establece el artículo 5 del RGPD, los datos son los adecuados para una determinada finalidad. En este conjunto de datos tenemos:

- Microdatos o identificadores directos de las personas: todas aquellas características que por sí mismas permiten la identificación de una persona.
- Identificadores indirectos: si bien no identifican a una persona, el cruce de varios identificadores indirectos podría permitir la identificación de una persona.

[4] Identificadores indirectos o cuasi-identificadores: variables que en combinación con otra información permiten la identificación de las personas, por ejemplo: género, estado civil, código postal u otros datos de localización, fechas significativas (nacimiento, fecha de ingreso hospitalario, etc.), profesiones, raza, pertenencia a grupos sociales minoritarios, ingresos económicos, etc.

- Datos especialmente protegidos o sensibles: los referidos en el artículo 9 del RGPD, datos financieros, datos de infracciones, etc.

El proceso de anonimización de los datos se realizará de forma estructurada teniendo en cuenta la finalidad que se pretende dar a los datos una vez anonimizados, garantizando la privacidad de las personas y evitando la distorsión de los resultados de la información anonimizada con respecto a los datos no anonimizados.

En esta fase del proceso deberá prestarse atención a las dificultades específicas de anonimización para determinadas variables como, por ejemplo, en el caso que fuera necesario anonimizar registros de voz, registros de imagen o información biométrica y/o genética.

3.7. ELIMINACIÓN/REDUCCIÓN DE VARIABLES

El objeto de esta fase 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 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.

3.8 SELECCIÓN DE LAS TÉCNICAS DE ANONIMIZACIÓN: CLAVES[5]

1. 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[6] 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.

   El algoritmo de hash permite que, partiendo de un mismo dato o microdato, podamos generar siempre la misma huella digital pero partiendo de una determinada huella digital nunca podremos obtener el dato original, garantizando la confidencialidad al tratarse de una operación matemática de un solo sentido. Las claves resultantes de la aplicación de un algoritmo de hash son comúnmente conocidas como "huella digital" por entender que representan de forma unívoca a un dato o microdato concreto.

   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

[5] Clave: Por clave se entenderá el resultado de aplicar cualquier técnica para garantizar la confidencialidad de los datos anonimizados. Por ejemplo, si la cadena AAA reemplaza al nombre de una persona AAA se convierte en una clave o seudónimo.
[6] Algoritmo de "hash": algoritmo matemático que permite generar la huella digital de un documento o unidad de información digital (archivo, imagen, cadena alfanumérica, etc.).

secreto de la huella digital que hemos generado. Una buena opción es el algoritmo HMAC basado en RFC2014[7] 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.

La utilización de HMAC en combinación con claves secretas no triviales y una política diligente de destrucción de claves puede servir para garantizar la irreversibilidad del proceso de anonimización. Cuando las claves utilizadas con HMAC se conservan pueden servir para generar datos seudonimizados que requieran una posterior reidentificación de los interesados. Mecanismos de hash con clave secreta pueden resultar útiles para enmascarar los datos. Sin embargo, deberá existir un procedimiento que permita la eliminación segura de las claves y la posibilidad de acreditar que el procedimiento se ha cumplido para garantizar la irreversibilidad del proceso.

2. 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.

3. 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.

4. 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.

[7] RFC2014: RFC ("Rule For Comments") es el acrónimo utilizado referido a las normas de facto que han dado lugar a estándares en Internet.

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 la clasificación de las variables de identificación**: es posible elaborar un procedimiento de forma que existan varios niveles de anonimización. Los datos identificativos que tienen el nivel de criticidad más bajo pueden ser anonimizados en la capa inicial con un determinado conjunto de claves secretas[8]. En esta fase se codifican todos los datos que permitan la identificación de las personas. A continuación, es posible proceder a una segunda anonimización de los datos personales con un segundo conjunto claves secretas con información que haya sido clasificada con un grado de criticidad mayor, y así sucesivamente pueden reanonimizarse los datos fortaleciendo de esta forma la cadena de anonimización tantas veces como sea necesario. Este método establecerá una dificultad proporcional a la criticidad o sensibilidad de la información anonimizada, multiplicando el esfuerzo necesario para la reidentificación de las personas.

- **Atendiendo a la organización interna:** en ocasiones el tratamiento de los datos se encuentra estructurado por departamentos dentro de una organización, de forma que cada departamento necesita acceder a la información con una finalidad distinta. En este caso, es posible utilizar conjuntos de claves secretas para proceder a la anonimización y reanonimización de forma específica para cada departamento. Este procedimiento de anonimización multicapa garantiza la privacidad de los datos anonimizados de forma interdepartamental, haciendo necesaria la participación de varios departamentos para llegar al dato original que únicamente estará a disposición del departamento que realizó la anonimización inicial.

[8] Ejemplo: La clave AAA reemplaza al nombre de una persona, sería posible aplicar un algoritmo de hash añadiendo texto aleatorio como AAA1k2j3j de manera que sobre el seudónimo AAA tendríamos una clave secreta (1k2j3j) que fortalecería la cadena de confidencialidad; en este caso se debería contemplar la posibilidad de disponer de una política de claves secretas para llevar a cabo la anonimización. Si sobre la cadena AAA aplicamos un algoritmo de hash MD5, la reidentificación sería relativamente sencilla por mecanismos de fuerza bruta, mientras que el hecho de añadir texto o una clave adicional al seudónimo implicaría un mayor esfuerzo para realizar la reidentificación de una persona.

- **Atendiendo a las garantías específicas de la política de anonimización del responsable del tratamiento de los datos anonimizados:** el receptor de la información anonimizada decide aplicar su política de anonimización para tratar la información con sus propias garantías de forma que en caso de brecha de reidentificación sea posible identificar el origen y la responsabilidad de la brecha en la anonimización.

En combinación con las técnicas de anonimización criptográfica pueden utilizarse otras técnicas como las que se describen a continuación:

5. PERTURBACIÓN DE DATOS: variación y supresión sistemática de datos que evita que las cifras resultantes faciliten información sobre casos específicos:

   - Microagregación: técnica utilizada para anonimizar datos numéricos consistente en la sustitución de valores numéricos concretos por el valor medio calculado para un determinado grupo de datos mediante la agrupación, segregación, supresión o sustitución de registros independientes.
   - Intercambio aleatorio de datos: introduce una distorsión aleatoria en un conjunto de microdatos manteniendo el detalle y estructura de la información original.
   - Datos sintéticos:
     - Distorsión de datos: se generan datos aleatorios que mantienen los resultados del conjunto de datos originales
     - Distorsión con microdatos híbridos: combinación de datos originales con datos sintéticos
   - Permutación de registros: intercambio de valores de datos con valor clave que garantiza valores promedios y distribuciones estadísticas.
   - Permutación temporal: movimiento aleatorio de rangos temporales que no genera distorsión sobre los resultados medios finales.
   - Redondeo: sustitución de variables por valores redondeados de forma aleatoria.
   - Reajuste de pesos: cuando se trabaja con muestras de datos conocidas, supone la distorsión de los valores de las muestras originales para evitar la reidentificación.
   - Ruido aleatorio: inyecta ruido manteniendo la estructura de datos originales.

6. REDUCCIÓN DE DATOS: se reduce el número de datos originales sin alterar los mismos, disminuyen el nivel de detalle de los datos originales evitando la presencia datos únicos o atípicos sin relevancia para el resultado final:

   - Eliminación de variables: eliminación de datos especialmente sensibles que pueden ser identificadores directos.
   - Reducción de registros: cuando tras aplicar otra medida los sujetos sigan siendo identificables.
   - Recodificación global: determinadas categorías de datos se agrupan en una nueva categoría reduciendo las posibilidades de reidentificación.
   - Codificación superior o inferior: para casos en los que valores superiores o inferiores de un rango sean identificables, consiste en ampliar o reducir el rango mayor o menor.

- Supresión de registros: eliminación de registros de datos que contienen datos que permiten la identificación de sujetos. Esta medida se utilizará cuando sea imposible anonimizar un determinado sujeto y se hará indicación expresa de los registros eliminados y el motivo por el que se excluyen del resultado final de la anonimización.

En relación a las técnicas de anonimización, el Grupo de Trabajo del artículo 29[9] emitió su Dictamen 05/2014 en el que se realiza un análisis técnico acerca de la robusted, 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. Esta distorsión es conocida como "diferencial de privacidad" que se tendrá en cuenta en los procesos de anonimización y cuyos índices deben ser cuantificables y asumibles dentro de la finalidad a la que vayan a destinarse los datos anonimizados, lo que puede significar cierto grado de desconfianza hacia los resultados de los análisis de la información.

El rápido avance tecnológico podría dejar obsoletas las técnicas de anonimización elegidas en un corto espacio temporal con la aparición de técnicas más seguras o la aparición de técnicas que permitan vulnerarlas. La selección de medidas técnicas de anonimización debería someterse también al análisis de riesgos mencionado para cada proceso de anonimización, ayudando al establecimiento de pautas o principios que permitan y ayuden al responsable del proceso de anonimización a optar o decidir por las técnicas más apropiadas en cada momento concreto.

3.9. SEGREGACIÓN DE LA INFORMACIÓN

En consonancia con los principios antes mencionados y con el fin de garantizar la confidencialidad de la información anonimizada, debe tenerse en cuenta la necesidad de elaborar un mapa de sistemas de información que garantice entornos separados para cada tratamiento de datos personales o información personal anonimizada.

Es recomendable que el proceso de anonimización se realice partiendo de los datos personales pero en un entorno segregado. A su vez, la explotación de la información anonimizada debe realizarse en un entorno ajeno a los entornos de explotación de los datos personales o al entorno en el que se realiza la anonimización de la información.

La segregación de entornos para los tratamientos de la información implicará también la segregación del personal que accede a la información y a los datos personales. Una garantía adicional para evitar la reidentificación es que aquellas personas implicadas en el tratamiento de la información personal anonimizada no tengan acceso a los datos personales no anonimizados o no puedan acceder al conocimiento de los mecanismos y claves de anonimización utilizados en los procesos de anonimización.

[9] El Grupo de Trabajo del artículo 29 (GT29) se establece como resultado de la aplicación de la Directiva europea de protección de datos (95/46/EC). Para más información acerca de su funcionamiento y estructura o consultar su Dictamen 5/2014 sobre técnicas de anonimización puede consultar este enlace.

3.10. PROYECTO PILOTO

Es recomendable la realización de un proyecto piloto con una pequeña muestra de datos de prueba (no reales) en el que puedan obtenerse de forma objetiva conclusiones acerca de la viabilidad de todas las propuestas de los miembros del equipo de anonimización. Algunos de los objetivos que pueden ser tenidos en cuenta para la realización del proyecto piloto son los siguientes:

- Arrojar luz sobre los resultados de las técnicas de anonimización propuestas.
- Comprobar la fortaleza de los resultados frente a posibles intentos de reidentificación.
- Cuantificar los costes del proceso de anonimización.
- Asegurar que los objetivos que pretenden los procedimientos de anonimización son viables.
- Garantizar el diferencial de privacidad[10] mediante pruebas encaminadas a evaluar y cuantificar las desviaciones o distorsiones resultantes del proceso de anonimización.
- Evaluar los resultados del proyecto frente a los resultados del análisis de riesgos.

Los resultados del proyecto piloto se deben compartir con el destinatario de la información anonimizada junto con los datos no reales que hayan sido utilizados para la realización del proyecto piloto, de tal forma que sea posible valorar si el resultado del proceso de anonimización se ajusta a los objetivos a los que se va a destinar la información y permitiendo a su legítimo destinatario valorar si el diferencial de privacidad es aceptable o si, por el contrario, la distorsión generada por el proceso de anonimización genera un diferencial de privacidad que hace que la información resultante no pueda ser utilizada con la finalidad que se persigue.

3.11. ANONIMIZACIÓN

Finalmente, en la fase de anonimización se realiza la disociación definitiva e irreversible de los datos personales. El proceso de anonimización deberá realizarse tantas veces como sea necesario según la finalidad de la información anonimizada y su destinatario. Es recomendable que cada uno de los destinatarios disponga de datos anonimizados con claves distintas al igual que es necesario que el proceso de anonimización se realice ad hoc con un objetivo, finalidad y destinatario concretos. En ningún caso se recomienda utilizar un proceso de anonimización de uso general con independencia del destinatario de la información, del tipo de información a anonimizar y la finalidad a la que se vayan a destinar los datos anonimizados.
A continuación se muestran algunas de las tareas o actividades que pueden realizarse durante la fase de anonimización:

[10] El diferencia de privacidad es la diferencia o distorsión resultante del análisis de la información personal anonimizada frente al análisis de los datos personales no anonimizados.

- Determinar la técnica de anonimización que sea más apropiada en función de las variables que hubieran sido identificadas en la fase de preanonimización.
- Planificación y asignación de tareas específicas a cada miembro del equipo de trabajo con relación a las funciones asignadas para cada perfil implicado en el proceso de anonimización.
- Determinar los recursos y equipo técnico necesarios para proceder a la anonimización de los datos.
- Validar la técnica de anonimización por expertos (unidad u organismo experto en estadística, ética,etc.)
- Aplicar la técnica seleccionada y ejecutar el proceso de anonimización; realizar pruebas.
- Ruptura relacional de las claves en función del uso de la información (uso interno y uso externo). Siempre que sea posible se utilizarán distintas claves en función del uso que vaya a darse a la información anonimizada.
- Recodificación o reducción de variables para los datos sensibles residuales tras el proceso de anonimización.
- Aplicar técnicas de reducción de datos (supresión de campos que no sean significativos para el uso posterior).
- Acotar el nivel de desagregación en función del nivel geográfico afectado por el fichero y la sensibilidad de la información.
- Aplicar técnicas de perturbación de los datos (modificar datos cuantitativos en pequeñas cantidades aleatorias, intercambiar atributos de forma controlada entre registros de zonas geográficas próximas, respetando las distribuciones).
- Validación y aprobación de los archivos anonimizados por expertos y por el equipo de evaluación.
- Revisión periódica del proceso.
- Auditar el proceso de anonimización y el uso posterior de los datos mediante métricas o escalas que proporcionen una interpretación objetiva de los resultados.

Durante el proceso de anonimización los datos biométricos, registros de voz o registros de imagen pueden presentar una complejidad específica que deberá abordarse en las fases iniciales del proceso de anonimización con carácter previo. Por ejemplo, en cuanto a los registros de voz cabe la posibilidad de realizar una transcripción previa con su respectiva eliminación de posibles identificadores (expresiones autóctonas, elementos epidícticos, identificadores retóricos, etc.) para posteriormente proceder a la reproducción de las transcripciones mediante dispositivos sintetizadores de voz en caso que fuera necesario mantener un registro sonoro.

Los registros de imágenes presentan su riesgo de reidentificación en el conjunto de la imagen, ya que en ocasiones puede reidentificarse a las personas por su entorno y no directamente por sus propios rasgos genéricos. Las variables de reidentificación de las personas mediante imágenes pueden ser múltiples, por lo que en ocasiones los datos de imagen requerirán un tratamiento específico para impedir la reidentificación de las personas. Por ejemplo, en el caso de una dolencia dermatológica determinada en el que la persona disponga de un determinado tatuaje o cicatriz que ponga de manifiesto su identidad, la imagen deberá someterse a un tratamiento digital que haga irreversible la reidentificación de la persona.

En cuanto a los datos biométricos, la finalidad de la información anonimizada puede suponer una limitación a la anonimización de la información, dando lugar a determinadas excepciones en las que los datos no puedan ser anonimizados a fin de evitar cualquier distorsión crítica que se pueda

producir con relación a la información no anonimizada. Estas situaciones se contemplarán en las fases iniciales de la anonimización y especialmente en la EIPD como riesgo implícito al propio proceso dadas las características de la información. En este caso, la propia EIPD puede plantear la necesidad de utilizar mecanismos de cifrado para el acceso a los datos biométricos de forma restringida y controlada, mecanismos que deberán ser acordados por el responsable del tratamiento y el responsable del tratamiento de la información anonimizada o cifrada.

4. FORMACIÓN E INFORMACIÓN AL PERSONAL IMPLICADO EN LOS PROCESOS DE ANONIMIZACIÓN Y AL PERSONAL QUE TRABAJA CON DATOS ANONIMIZADOS

El personal implicado en el proceso de anonimización debe cumplir todos los requisitos de formación e información relativos al cumplimiento de la normativa de protección de datos personales, especialmente en lo relativo a las medidas de seguridad de índole técnica y organizativas a las que se refiere el artículo 32 del RGPD.

Una vez que los datos personales hayan sido anonimizados, el personal con acceso a la información anonimizada será también informado de:

- La existencia y aplicación de la política de anonimización.
  - Principios de calidad en el diseño de los procesos de anonimización.
  - Objetivos fijados en la gestión de riesgos (EIPD).
  - Estructura y responsabilidades del equipo de trabajo implicado en los procesos de anonimización.
  - Objetivos y finalidad de la información anonimizada.
  - Variables de anonimización: identificación y clasificación.
  - Técnicas de anonimización utilizadas.
- Términos de uso y acceso a la información anonimizada.
- Medidas de control del personal con acceso a la información anonimizada (trazabilidad).
- Obligaciones y deberes en caso de ruptura de la cadena de anonimización[11] que haga posible la reidentificación de los interesados.

La información facilitada al personal se proporcionará de forma tal que pueda ser auditable, es decir, existirá un registro con la información proporcionada y al personal que le fue facilitada.

5. GARANTIAS

El proceso de anonimización no puede asegurar la imposibilidad de reidentificación de las personas en términos absolutos, motivo por el cual se deben de tener en cuenta las garantías jurídicas necesarias para preservar los derechos de los interesados.

[11] Cadena de confidencialidad o anonimización: suma de operaciones realizadas con el fin de eliminar variables de identificación directa o microdatos y datos de identificación indirecta.

En este sentido algunos de los aspectos que deben ser tenidos en cuenta 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.

Estas y otras posibles garantías que pudieran ser necesarias para el tratamiento de información anonimizada serán tenidos en cuenta en la EIPD como parte de las salvaguardas encaminadas a minimizar los daños ante una eventual reidentificación de los interesados.

6. AUDITORÍA DEL PROCESO DE ANONIMIZACIÓN

El objeto de la auditoría del proceso de anonimización es garantizar el cumplimiento de la política de anonimización, proporcionando una opinión objetiva sobre el conjunto del proceso de anonimización. La auditoría puede ser interna o externa y tendrá carácter periódico.

La calidad de la propia auditoría es fundamental para el mantenimiento de la confianza de los interesados en los procesos de anonimización, ya que la falta de confianza de los interesados en la confidencialidad de los procesos de anonimización podría provocar inquietud social repercutiendo negativamente en la explotación de datos anonimizados.

Los resultados de la auditoría pueden darse a conocer a los interesados facilitándoles información sobre las probabilidades de reidentificación y las buenas prácticas acreditadas en el proceso de anonimización. Con el fin de garantizar la calidad de la auditoría es recomendable el uso de normas, metodologías y estándares internacionales de reconocido prestigio.

La auditoría del proceso de anonimización mostrará resultados relativos a los objetivos de calidad de los procesos de anonimización que inicialmente hubieran sido previstos por el responsable del tratamiento.

El responsable del tratamiento o el responsable del tratamiento de los datos anonimizados velará por la existencia de informes periódicos de auditoría en los que al menos se haga constar:

- El alcance y objetivo de la auditoría.
- Definición del equipo auditor y recursos utilizados en la realización de la auditoría.
- Fases y planificación de la auditoría.

- Pruebas y verificaciones realizadas.
- Valoración de los resultados.
- Propuestas para la mejora del proceso de anonimización.
- Auditoría de la explotación de la información anonimizada.

La auditoría conllevará las comprobaciones necesarias encaminadas a verificar la implantación de las propuestas de mejora del proceso de anonimización y permitirá la verificación y el seguimiento de la eficacia de las medidas implantadas.

7. DOCUMENTACIÓN

La política de anonimización aplicable deberá estar documentada y accesible al personal implicado en el tratamiento de datos anonimizados. A continuación se muestra un posible esquema del contenido documental que puede ser tenido en cuenta para el proceso de anonimización:

- Política de uso y acceso a los datos anonimizados: obligaciones del personal.
- Documento de aplicabilidad de medidas de anonimización que contendrá al menos:
  - Responsables del proceso de preanonimización y anonimización.
  - Medidas organizativas.
  - Definición de variables de identificación.
  - Mecanismos técnicos de anonimización.
  - Política de claves
  - Acuerdos de confidencialidad.
- Normas y procedimientos.
- Informes y dictámenes:
  - Del equipo de viabilidad si hubiera sido definido.
  - Del equipo de seguridad.
  - De análisis de riesgos (EIPD).
  - De auditoría de la información y el proceso de anonimización.

El valor de la gestión documental que se realice del proceso de anonimización tiene especial relevancia en tanto que refleja de forma justificada las actuaciones que el responsable del tratamiento y el responsable del tratamiento de la información anonimizada realizan con el fin de proteger la privacidad y los datos de los interesados.

La documentación se actualizará siempre que sea necesario por cambios en el proceso de anonimización, en los requisitos legales o por condicionantes de la evolución tecnológica.

8. CONCLUSIÓN

Los procesos de anonimización y seudonimización son una herramienta válida para garantizar la privacidad de los datos personales y sus limitaciones son inherentes al avance de la tecnología.

Existe una proporcionalidad manifiesta en lo que respecta a la capacidad tecnológica de anonimizar y la posibilidad de la reidentificación de las personas cuyos datos han sido

anonimizados, es decir, la misma capacidad de la tecnología para anonimizar datos personales puede ser utilizada para la reidentificación de las personas. Además, hay que tener en cuenta el riesgo que la propia sociedad de la información añade a los datos anonimizados, riesgo que por otra parte evoluciona a lo largo del tiempo, por lo que habrá que contemplar el riesgo de los procesos de anonimización como una contingencia latente a lo largo de la vida de la información y no en un momento concreto, y, en consecuencia, las medidas encaminadas a valorar y gestionar los riesgos deben tener carácter periódico.

No es posible considerar que los procesos de anonimización garanticen al 100% la no reidentificación de las personas, por lo que será necesario sustentar la fortaleza de la anonimización en medidas de evaluación de impacto (EIPD), organizativas, de seguridad de la información, tecnológicas y, en definitiva, cualquier medida que sirva tanto para atenuar los riesgos de reidentificación de las personas como para paliar las consecuencias de que éstos se materialicen.

9. REFERENCIAS

- "Dictamen 05/2014"del Grupo de Trabajo del Artículo 29 sobre técnicas de anonimización.
- "Dictamen 06/2014" del Grupo de Trabajo del Artículo 29 sobre la noción de interés legítimo a la que se refiere el artículo 7 de la Directiva 95/46/EC.
- "Código de buenas prácticas de las estadísticas europeas para los servicios estadísticos nacionales y comunitarios", adoptado por el Comité del Sistema Estadístico Europeo el 28 de septiembre de 2011 (EUROSTAT).
- "Looking Forward: De-identification Developments – New Tools, New Chalenges" (May 2013, Information & Privacy Commissioner Ontario, Canada).
- "De-identification Protocols: Essential for Protecting Privacy", (June 2014, Information & Privacy Commissioner Ontario, Canada).
- "Dispelling the Myths Surrounding De-identification: Anonymization Remains a Strong Tool for Protecting Privacy", (June 2011, Information & Privacy Commissioner Ontario, Canada).
- "Big Data and Innovation, Setting the Record Straight: De-identification Does work" (June 2014, Information & Privacy Commissioner Ontario, Canada).
- "Pan-Canadian De-Identification Guidelines for Personal Health Information", (2007, Information & Privacy Commissioner Ontario, Canada).
- "Anonymisation: managing data protection risk", (November 2012, Information Commissioner´s Office, UK).
- "CNIL – guide sécurité des données", (2010, Commission nationale de l'informatique et des libertés).
- "Lineamientos para la anonimización de microdatos", (Agosto 2014, Dirección de Regulación, Planeación, Estandarización y Normalización –DIRPEN- Colombia).
- "Norma PNE 178301, Ciudades Inteligentes. Datos abiertos (Open Data) – Versión para Información Pública".
- "A Systematic Review of Re-Identification Attacks on Health Data", (US National Library of Medicine, National Institutes of Health).
- "Perspectives on Heal Data De-identification" (Privacy Analitycs, Khaled EI emam, PhD).

AGENCIA ESPAÑOLA DE PROTECCIÓN DE DATOS