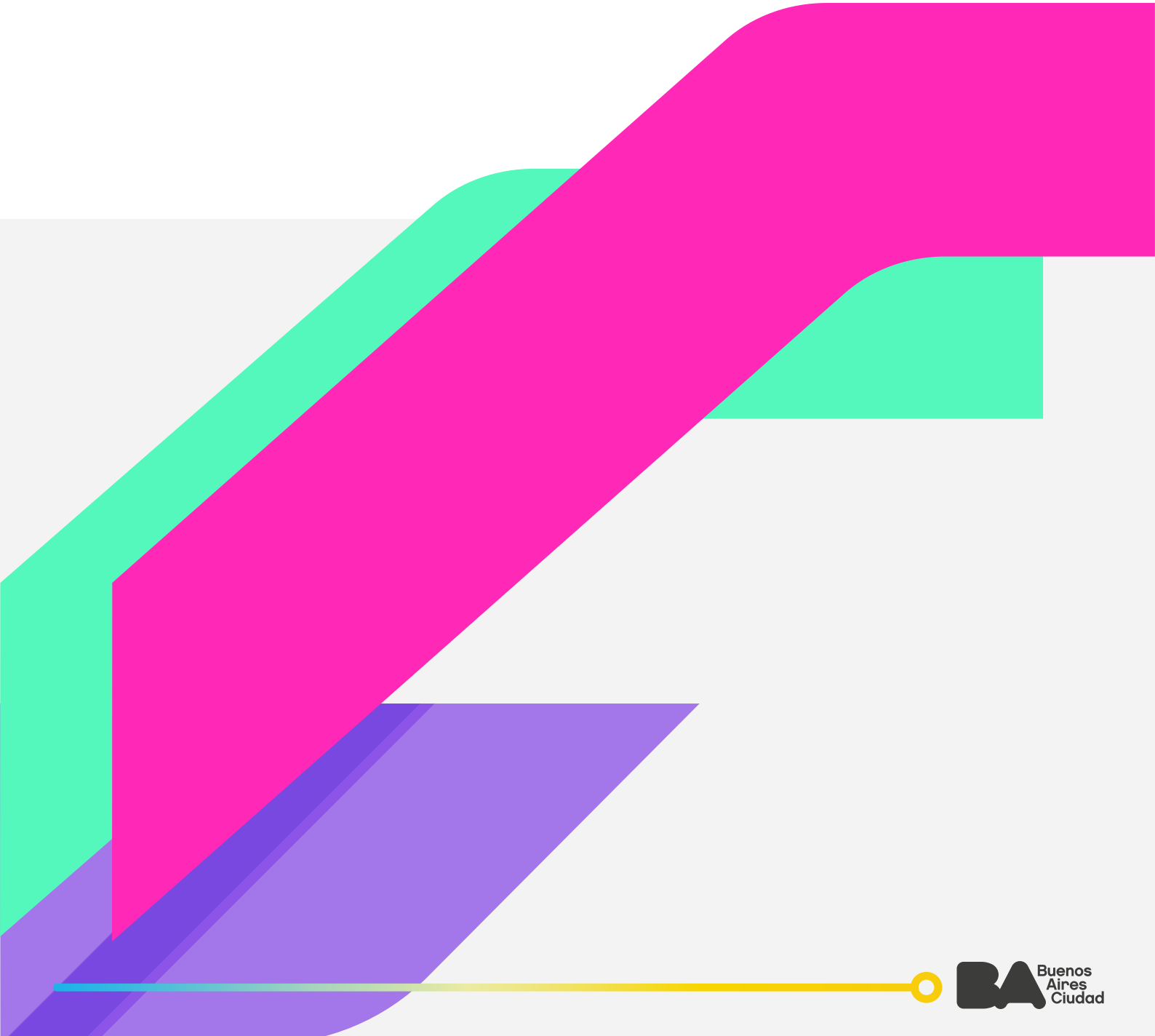
Gobernanza de Datos

Consejos y recomenadaciones para la **anonimización de los datos personales**

SECRETARÍA DE INNOVACIÓN Y TRANSFORMACIÓN DIGITAL
SUBSECRETARÍA DE POLÍTICAS PÚBLICAS BASADAS EN EVIDENCIA



Buenos Aires Ciudad

Jefe de Gobierno
Horacio Rodríguez Larreta
Jefe de Gabinete
Felipe Miguel
Secretario de Innovación y Transformación Digital
Diego Fernández
Subsecretaria de Políticas Públicas Basadas en Evidencia
Melisa Breda

Índice

1. Anonimización

¿Qué es la anonimización?

La anonimización, tal como define la Red Iberoamericana de Protección de Datos, es la "aplicación de medidas de cualquier naturaleza dirigidas a impedir la identificación o reidentificación de una persona física sin esfuerzos desproporcionados".

En la misma línea, la Agencia Española de Protección de Datos Personales en su guía "Orientaciones y garantías en los procedimientos de anonimización de datos personales" del año 2016, la define como "aquel proceso que tiende a 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 conlleve una distorsión de los datos reales". Un análisis masivo de los datos que puedan derivar de los datos anonimizados no debería diferir del análisis que pudiera obtenerse si hubiera sido realizado con datos no anonimizados.

La Ley Nacional N° 25.326 de protección de los datos personales, en su artículo N°2 define a la **disociación de datos** como "todo aquel tratamiento donde la información obtenida no pueda asociarse a persona determinada o determinable".

En tal sentido, la Unión Europea, a través del Reglamento General de Protección de Datos (RGPD), señala que 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"**. 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.

El riesgo residual de reidentificación es la probabilidad de que, a través de técnicas de conversión de los datos, se pueda individualizar a la persona. Esta situación podría darse en los siguientes casos:

- Si se relacionan datos que no son anónimos, se podría llegar a la identidad de alguien.
- Si se desarrollan tecnologías que permitan descifrar la anonimización.
- Si algunos datos dejan de ser anónimos.

Cualquier proceso sólido de anonimización debe evaluar el riesgo de reidentificación, que debe gestionarse y controlarse a lo largo del tiempo.

Cabe mencionar que el riesgo de reidentificación nunca puede considerarse nulo, pero, en cualquier caso, la anonimización ofrece mayores garantías de privacidad a las personas.

La presente *guía* tiene como objetivo ofrecer **criterios indicadores y orientadores** en el ámbito del Poder Ejecutivo del Gobierno de la Ciudad Autónoma de Buenos Aires **sobre buenas prácticas en materia de la anonimización de datos personales.**

2. Proceso

Los datos se han convertido en un elemento clave para casi cualquier proceso de nuestra vida cotidiana. Cada vez hay más formas de recoger datos, y mayor capacidad para procesarlos y compartirlos.

Por eso, es crítico poder garantizar la privacidad de las personas usuarias y la protección de sus datos personales, entendidos como derechos fundamentales. Realizar un proceso de disociación antes de disponibilizar el dato en cuestión es una buena práctica a seguir.

Como señalamos previamente, el proceso de anonimización posee dos objetivos principales:

1. Evitar la individualización de personas en un conjunto de datos.
2. Eliminar o reducir al mínimo los riesgos de reidentificación de esos datos anonimizados, manteniendo la veracidad de los resultados del tratamiento de los mismos.

En el proceso de anonimización se deberá producir la ruptura de la **cadena de identificación de las personas.** Esta cadena se compone por:

- **Microdatos o datos** que permiten la **identificación directa de una persona, como** por ejemplo, el DNI.
- **Datos de identificación indirecta** que pueden permitir la reidentificación, como por ejemplo, información de otras bases de datos, redes sociales, buscadores, blogs, etc.

La Identificación indirecta es aquella que puede tener lugar como consecuencia de la combinación de información proveniente de varias fuentes, que puede facilitar la reidentificación de las personas, a pesar de que sus datos hayan 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.

Durante el proceso de anonimización será necesario prever con un equipo de trabajo las consecuencias de una eventual reidentificación que pudiera generar un perjuicio o merma en sus derechos.

3. Equipo de trabajo

Para llevar adelante la anonimización, se recomienda, en primer lugar, la creación de un equipo de trabajo para que exista una división de tareas y responsabilidades, y un análisis del riesgo adecuado.

Es recomendable realizar la división de funciones teniendo en cuenta los roles definidos en la guía de roles y responsabilidades.

4. Reidentificación

Los ataques contra la anonimización pueden materializarse en forma de intentos deliberados o involuntarios de reidentificación, brechas de seguridad o divulgación de datos al público.

La reidentificación de los datos personales siempre significa un riesgo en la vida privada de una persona. 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.

De todos modos, corresponde que el equipo de trabajo realice un análisis del riesgo en forma previa a la publicación o transferencia de los datos anónimos, el cual se desarrollará en los puntos subsiguientes.

4.1. Análisis de riesgo

Antes de publicar o transferir datos anónimos, en el diseño y la realización de un sistema de información, corresponde que el equipo de trabajo realice un análisis de riesgos de reidentificación. A continuación se muestran algunas de las fases o etapas que deberían ser tenidas en cuenta:

4.1.1. Identificación y categorización de datos implicados en el proceso de anonimización

Es imprescindible determinar los datos implicados en el proceso de anonimización, con el fin de poder realizar la valoración de los riesgos inherentes a cada uno de ellos.

Algunos de los puntos que debemos tener en cuenta en la elaboración del análisis de riesgos son:

- Datos personales a anonimizar.
- Variables de identificación asociadas.
- Procesos de anonimización a utilizar.
- Sistemas de información implicados: hardware utilizado, limitación del software de anonimización, en relación a los datos que sea preciso anonimizar.

4.1.2. Identificación de riesgos

Es la catalogación inicial de riesgos atendiendo a tres categorías:

- Riesgos de reidentificación existentes conocidos.
- Riesgos potenciales de reidentificación.
- Riesgos no conocidos.

Algunos de los riesgos que deben ser tenidos en cuenta son:

- Riesgos de reidentificación por correlación con otros conjuntos de datos.
- 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.
- Existencia de un atacante o adversario potencial que asume el rol de “perseguidor”.

- Existencia de un sujeto que conoce la identidad de una persona en un bloque de información y que pretende obtener mayor información, etc.

Para la evaluación de cada riesgo, como buena práctica, se recomienda consultar el apartado nº 33 de la guía "Orientaciones y garantías en los procedimientos de anonimización de datos personales" del año 2016, de la Agencia Española de Protección de Datos, título N° 3.3. "EVALUACIÓN DE RIESGOS DE REIDENTIFICACIÓN".

4.1.3. Informe de riesgos

Finalmente, consideramos apropiado que un equipo técnico-legal realice una evaluación a nivel ejecutivo y refleje, en un informe, los riesgos existentes, y las técnicas a implementar para la anonimización. Dicho informe deberá contemplar los umbrales de riesgos aceptables para cada proceso de anonimización.

5. Selección de técnicas de anonimización

Es incuestionable la utilidad que tienen los algoritmos de cifrado cuando necesitamos anonimizar datos. Desde la Subsecretaría de Políticas Públicas Basadas en Evidencia, sugerimos implementar los siguientes métodos de anonimización, dependiendo de la magnitud de los datos:

5.1. Algoritmos de HASH

Recomendado para datos de texto simple (por ejemplo, nombre, género, nacionalidad, etc).

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. Este algoritmo de cifrado es de gran utilidad para anonimizar datos de texto simple.

Al utilizar un algoritmo de HASH (como por ejemplo, SHA-1) para ocultar o anonimizar datos, se genera una clave o huella digital que puede utilizarse para reemplazar el dato real, y dificultar su reconstrucción. Cualquier variación en el dato original dará lugar a una huella digital diferente: en términos informáticos, podría decirse que la modificación de un solo bit en la información original almacenada daría lugar a una clave o huella digital distinta.

El algoritmo de HASH permite que, partiendo de un mismo dato, podamos generar siempre la misma huella digital; pero, partiendo de una determinada huella digital, nunca podremos obtener el dato original. Esto garantiza la confidencialidad, al tratarse de una operación matemática de un solo sentido.

Sin embargo, un algoritmo de HASH por sí solo no es suficiente para lograr la anonimización, ya que pequeñas cadenas de texto (por ejemplo, microdatos correspondientes al código postal de una persona, un número de teléfono, etc), pueden ser fácilmente reidentificados 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 la confidencialidad de la huella digital que hemos generado.

Los mecanismos de HASH con clave confidencial pueden resultar útiles para enmascarar los datos. Sin embargo, deberá existir un procedimiento que permita la eliminación segura de las claves, y posibilite acreditar el cumplimiento del procedimiento, para garantizar su irreversibilidad.

Para mayor detalle sobre la utilización de técnicas de HASH, se recomienda la lectura de la guía "INTRODUCCIÓN AL HASH COMO TÉCNICA DE SEUDONIMIZACIÓN DE DATOS PERSONALES" de la Agencia Española de Protección de Datos Personales.

5.2. Algoritmos de HMAC

Recomendado para microdatos (por ejemplo: código postal, año, edad, etc.).

Un algoritmo HMAC es una construcción específica para calcular un código de autentificación de mensaje (en inglés, Message Authentication Code, "MAC") que implica una función HASH criptográfica en combinación con una llave criptográfica confidencial. Es recomendable utilizar este algoritmo para anonimizar microdatos, ya que sobre la huella digital o clave resultante del algoritmo de HASH se podrá aplicar un algoritmo criptográfico que genere una nueva clave o huella digital o clave, en función de una clave confidencial.

La utilización de HMAC en combinación con claves secretas no triviales y una política diligente de destrucción de claves, puede garantizar la irreversibilidad del proceso de anonimización.

6. Posibles técnicas complementarias

6.1. Reducción de datos

La reducción de los datos a utilizar minimiza el número de datos originales y su nivel de detalle, sin alterar los mismos. Esto evita la presencia de datos únicos o atípicos sin relevancia para el resultado final:

6.1.1. Eliminación de variables: eliminación de datos especialmente sensibles que pueden ser identificadores directos.

6.1.2 Reducción de registros: cuando, tras aplicar otra medida, los sujetos sigan siendo identificables.

6.1.3. Recodificación global: determinadas categorías de datos se agrupan en una nueva categoría, reduciendo las posibilidades de reidentificación.

6.1.4. Codificación superior o inferior: para casos en los que valores superiores o inferiores de un rango sean identificables, este consiste en ampliar o reducir el rango.

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

6.2. Sello de tiempo

Consideramos buena práctica utilizar algoritmos de sello de tiempo, con el fin de dejar asentada la fecha y hora de la anonimización. También pueden utilizarse algoritmos de firma electrónica, que permitan identificar a quien haya realizado la anonimización.

6.3. Datos biométricos, voz e imagen

Los datos biométricos, registros de voz o de imagen pueden presentar complejidades específicas durante el proceso de anonimización. Las mismas deberán abordarse en las fases iniciales del proceso.

6.3.1. Registros de voz: es posible realizar una transcripción previa sin incluir posibles identificadores (por ejemplo, expresiones autóctonas, identificadores retóricos, etc) para reproducirla mediante dispositivos sintetizadores de voz, en caso que fuera necesario mantener un registro sonoro.

6.3.2. Registros de imagen: las imágenes presentan riesgos por múltiples variables , ya que en ocasiones puede reidentificarse a las personas por su entorno o por sus rasgos particulares. Es por ello que, en ocasiones, los datos de imagen requerirán un tratamiento específico para impedir la reidentificación.

Por ejemplo, en caso de que una persona presente un determinado tatuaje o cicatriz que ponga de manifiesto su identidad, la imagen deberá someterse a un tratamiento digital que impida su reidentificación.

6.3.3. Datos biométricos: podemos encontrar determinadas excepciones en las que los datos no podrán ser anonimizados, a fin de evitar cualquier distorsión crítica que se pueda producir con relación a la información original. Esto puede ocurrir cuando la finalidad de la información anonimizada limita la anonimización de la información.

Finalmente, para profundizar más sobre técnicas de anonimización, se recomienda la lectura de las guías y orientaciones referidas a continuación:

- Red Iberoamericana de Protección de Datos Personales: Estándares de Protección de Datos Personales.
- Guía de la Agencia Española de Protección de Datos Personales: Malentendidos en la anonimización.
- Guía de la Agencia Española de Protección de Datos Personales: Orientaciones procedimientos de anonimización.

- 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.
- Resolución 47/2018 - ASI "MEDIDAS DE SEGURIDAD RECOMENDADAS PARA EL TRATAMIENTO Y CONSERVACIÓN DE LOS DATOS PERSONALES EN MEDIOS INFORMATIZADOS

7. Anexo

7.1. Relaciones

- Guía de clasificación de datos.
- Guía de roles de gobernanza de datos.

7.2. Bibliografía consultada

- Ley de Protección de Datos Personales Nacional N° 25326
- Ley de Protección de Datos Personales de la Ciudad Autónoma de Buenos Aires N° 1845
- AEPD - Malentendidos en la anonimización
- AEPD- GUIA "Orientaciones y garantías en los procedimientos de anonimización de datos personales" del año 2016, de la Agencia Española de Protección de Datos,
- 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).
- Resolución 47/2018 - ASI "MEDIDAS DE SEGURIDAD RECOMENDADAS PARA EL TRATAMIENTO Y CONSERVACIÓN DE LOS DATOS PERSONALES EN MEDIOS INFORMATIZADOS"
- AEPD - INTRODUCCIÓN AL HASH COMO TÉCNICA DE SEUDONIMIZACIÓN DE DATOS PERSONALES
- Estándares de Protección de Datos Personales - Red Iberoamericana de Protección de Datos Personales
- Reglamento General de Protección de Datos - Unión Europea

8. Contacto

Ante cualquier duda o comentario sobre este documento, podés escribirnos a **datosgcba@buenosaires.gob.ar**