SECRETARÍA
gob digital

Guía introductoria a la anonimización de datos

Introducción

En las últimas décadas, hemos sido testigos de un desarrollo vertiginoso en las tecnologías digitales que ha transformado radicalmente nuestra forma de comunicarnos, hacer negocios y mejorar los servicios públicos. Este progreso tecnológico ha convertido a los datos, en un recurso invaluable para el desarrollo e innovación en múltiples sectores. No obstante, este avance trae consigo desafíos significativos, especialmente en lo que respecta a la privacidad y protección de datos. En Chile, la actualización legislativa y la demanda de prácticas de privacidad más rigurosas enfatizan la necesidad de abordar estos retos de manera efectiva. La anonimización de datos se erige como una herramienta esencial no sólo para proteger la privacidad de los individuos, sino también para permitir que las organizaciones aprovechen el valor de los datos recolectados, asegurando la protección de la privacidad y los datos personales de los usuarios.

El uso de estos datos no sólo impacta los resultados de empresas privadas, sino que también influye en la mejora de los servicios públicos y en la toma de decisiones, además de contribuir a la realización de estudios sociales, científicos y económicos. Tanto en el caso de los datos abiertos, como en el manejo general de datos, es fundamental garantizar la privacidad de los usuarios y usuarias, y la protección de sus datos personales.

El objetivo de este documento es ofrecer una introducción básica y concisa, orientada principalmente a entidades y organizaciones que gestionan datos personales. Este tratamiento abarca desde la obtención hasta la publicación de información, independientemente del soporte o medio. Sin embargo, es importante aclarar que, por cuestiones de alcance, este documento se centra exclusivamente en la anonimización de bases de datos estructuradas y no pretende ser un manual exhaustivo. Más bien, es una primera aproximación para comprender los conceptos, riesgos y técnicas disponibles, así como la complejidad inherente a cualquier proceso de anonimización de datos.

La elaboración de esta guía se ha basado apoyado en diversas fuentes autorizadas, incluyendo la "Guía de Anonimización" de la AEPD, el "Informe de Anonimización" del Ministerio de Asuntos Económicos y Transformación Digital, los "Criterios de Disociación de Datos Personales" de AGESIC, entre otros. Estos documentos proporcionan una base sólida para los métodos y herramientas discutidos, asegurando una comprensión integral y aplicada de las técnicas de anonimización.

Todo lo anterior, sin perjuicio de los procedimientos de anonimización de datos personales que se establezcan conforme a los reglamentos dictados en virtud del Proyecto de Ley que regula la Protección y el Tratamiento de los Datos Personales y crea la Agencia de Protección de Datos Personales.

1. Definiciones:

- Anonimización: procedimiento irreversible en virtud del cual un dato personal no puede vincularse o asociarse a una persona determinada, ni permitir su identificación, por haberse destruido o eliminado el nexo con la información que vincula, asocia o identifica a esa persona. Un dato anonimizado deja de ser un dato personal. (PDL)
- Atributos objetivos: son aquellos atributos de datos que se pretenden analizar y que son esenciales para el propósito del procesamiento de datos.
- Clase de Equivalencia: grupo de registros en una base de datos modificados para contener al menos 'k' registros idénticos, evitando que cualquier registro pueda ser directamente vinculado a un individuo. Protege contra la reidentificación al hacer que cada registro sea indistinguible de al menos 'k-1' otros registros en el mismo conjunto.
- Datos de carácter personal o datos personales: cualquier información vinculada o referida a una persona natural identificada o identificable. Se considerará identificable toda persona cuya identidad pueda determinarse, directa o indirectamente, en particular mediante uno o más identificadores, tales como el nombre, el número de cédula de identidad, el análisis de elementos propios de la identidad física, fisiológica, genética, psíquica, económica, cultural o social de dicha persona, excluyendo aquellos casos en que el esfuerzo de identificación sea desproporcionado. (PDL)
- Datos sensibles: sólo tendrán esta condición aquellos datos personales que revelen el origen étnico o racial, la afiliación política, sindical o gremial, hábitos personales, las convicciones ideológicas o filosóficas, las creencias religiosas, los datos relativos a la salud, al perfil biológico humano, los datos biométricos, y la información relativa a la vida sexual, a la orientación sexual y a la identidad de género de una persona natural. (PDL)
- Dato abierto: un dato digital con las características técnicas y jurídicas necesarias para que pueda ser usado, reutilizado y redistribuido libremente por cualquier persona u órgano de la Administración del Estado, en cualquier momento y lugar.
- Datos desidentificados: son aquellos datos que han sido procesados para eliminar o modificar los identificadores personales directos, de tal manera que la posibilidad de asociar los datos con una persona original se minimiza, pero no se elimina completamente, lo que implica que podrían ser reidentificados bajo ciertas condiciones.
- Datos anonimizados: son datos que permitían identificar a una persona física o jurídica en su forma original, pero que han pasado por un proceso de anonimización que imposibilita la reidentificación del propietario, por lo que ya no son datos personales.
- Datos seudonimizados: tratamiento realizado sobre datos personales de manera que ya no puedan atribuirse a un individuo (interesado), a menos que se emplee información adicional, y 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 atribuyen a una persona física identificada o identificable.

- Dato sintético: datos generados que no tienen una correspondencia directa con datos reales, pero son creados específicamente para su uso en análisis y procesos sin comprometer la privacidad.
- Desidentificar: consiste en la eliminación de identificadores (por ejemplo, nombre, dirección, número de documento nacional de identidad) que permitan reconocer directamente a un individuo. La desidentificación a veces se entiende erróneamente con la anonimización, sin embargo, es solo el primer paso de la anonimización.
- Identificación: proceso de reconocimiento o determinación de la identidad específica de una persona usando información o datos únicos asociados a ella. En el contexto de manejo de datos, se refiere a la capacidad de asociar información disponible con una persona física conocida o identificable.
- Identificadores directos: son datos que identifican de manera unívoca a una persona, como nombres, números de identificación personal o direcciones exactas.
- Identificadores indirectos o seudoidentificadores: son aquellos datos que al combinarse con la misma o diferentes fuentes de datos pueden permitir identificar a un individuo (datos sociodemográficos, configuración del navegador, ubicación geográfica, etc). También se conocen como cuasi-identificadores.
- Minimización de Datos: proceso que limita la recolección, almacenamiento y uso de datos personales a lo estrictamente necesario para cumplir con el propósito especificado. Este enfoque reduce el riesgo de violaciones de privacidad, asegurando que no se recopilen ni retengan más datos de los necesarios.
- Persona identificada: toda persona cuya identidad ya se conoce. Por ejemplo, “Juan Pérez”.
- Persona identificable: cualquier persona cuya identidad sea desconocida, pero pueda determinarse, directa o indirectamente, mediante uno o más identificadores.
- Reidentificación: identificar a las personas específicas a las que pertenecen los datos a partir de ellos. Es uno de los riesgos clave a mitigar en un proceso de anonimización de datos.
- Registro: una colección de atributos relacionados con un individuo, que forman una unidad dentro de un conjunto de datos.
- Seudonimización: tratamiento de datos personales que se efectúa de manera tal que ya no puedan atribuirse a un titular 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 natural identificada o identificable. (PDL)
- Tratamiento de datos: cualquier operación o complejo de operaciones o procedimientos técnicos, de carácter automatizado o no, que permitan recolectar, almacenar, grabar, organizar, elaborar, seleccionar, extraer, confrontar, interconectar, disociar, comunicar, ceder, transferir, transmitir o cancelar datos de carácter personal, o utilizarlos en cualquier otra forma.

2. Anonimización

2.1. Qué es la anonimización

La anonimización es una técnica esencial en la gestión de datos personales, que busca prevenir eficazmente tanto la identificación como la reidentificación de individuos. Este proceso integra medidas técnicas y organizativas para minimizar los riesgos de reidentificación y generalmente implica la eliminación de identificadores directos, así como la modificación de identificadores indirectos, asegurando al mismo tiempo la integridad y utilidad de los datos tratados. Conforme al Reglamento General de Protección de Datos (GDPR) de la Unión Europea, la anonimización debe asegurar que el individuo sea irreconocible, equilibrando costos y tecnología disponible para impedir identificaciones, aun frente al avance tecnológico.

A pesar de su objetivo de eliminar totalmente el riesgo de reidentificación, prácticamente siempre subsiste un riesgo residual. Esto se debe a la posibilidad de combinar datos parcialmente anonimizados y al desarrollo de nuevas tecnologías que podrían desentrañar métodos de anonimización anteriores. Errores o modificaciones en el manejo de datos también pueden hacer que se vuelvan identificables.

Por ello, es crucial implementar un proceso de anonimización sólido y dinámico que incluya evaluaciones de riesgo constantes y adaptación de las medidas de seguridad, para mantener una protección efectiva y duradera de la privacidad.

2.2. Casos en los que se aplica la anonimización

La anonimización no se limita solo a los datos abiertos o a cumplir con requisitos de transparencia; su aplicación es fundamental en diversos contextos donde se procesan datos personales. A continuación se describen algunos casos donde se aplican técnicas de anonimización:

2.2.1. Publicación de datos abiertos:

Anonimizar datos para publicaciones abiertas permite a las organizaciones compartir información valiosa sin comprometer la privacidad de las personas.

2.2.2. Compartir datos entre organizaciones:

Crucial para proteger los datos personales durante colaboraciones o servicios tercerizados, asegurando que estos datos estén seguros antes de su transferencia.

2.2.3. Intercambio interno de datos (datos anonimizados):

las organizaciones podrían considerar utilizar datos anonimizados para su intercambio interno en los siguientes casos:

- El intercambio interno de datos no requiere datos personales detallados (por ejemplo, para el análisis de tendencias).
- Los datos involucrados son de naturaleza más sensible (por ejemplo, información financiera).

2.3. Identificación y clasificación de los datos.

Un registro de datos personales se compone de atributos de datos que tienen diversos grados de identificabilidad y sensibilidad a un individuo. La tabla y los ejemplos siguientes ilustran cómo un atributo de datos se clasifica normalmente dentro de un registro de datos.

	Identificadores directos	Identificadores indirectos o seudoidentificadores	Atributos objetivos
Definición	Son atributos de datos que son exclusivos de un individuo y que pueden identificarlo de manera unívoca y directa.	Son atributos de datos que no son exclusivos de un individuo, pero pueden permitir identificar a un individuo cuando se combinan con otra información.	Estos atributos representan la finalidad principal para la que se recolectaron los datos y son cruciales para el análisis o las operaciones para las cuales fueron recogidos. Estos atributos, como los ingresos anuales o el diagnóstico médico, pueden incluir información sensible y podrían resultar en consecuencias adversas si se divulgan, ya que podrían afectar la privacidad o seguridad de la persona involucrada.
Ejemplos comunes en un conjunto de datos	• Nombre •Dirección de correo electrónico • Número de teléfono móvil • Número de RUT • Número de pasaporte • Número de certificado de nacimiento • Nombre de usuario	• Edad • Género • Carrera • Fecha de nacimiento • Dirección • Código postal • Dirección del trabajo • Nombre de la empresa • Estado civil	• Origen étnico • Afiliación política, sindical o gremial. • Situación socioeconómica. • Las convicciones ideológicas o filosóficas • Creencias religiosas

	de redes sociales	• Altura • Peso • Dirección de protocolo de Internet (IP) • Número de matrícula del vehículo • Número de bastidor del vehículo • Localización del Sistema de Posicionamiento Global (GPS)	• Datos relativos a la salud y al perfil biológico humano • Datos biométricos • Información relativa a la vida sexual, a la orientación sexual y a la identidad de género. • Transacciones (por ejemplo, compras) • Sueldo • Calificación crediticia • Póliza de seguro • Diagnóstico médico • Estado de vacunación

Cualquier atributo que no sea esencial para el propósito del procesamiento debe ser eliminado (minimización de datos) para reducir los riesgos asociados al manejo de datos innecesarios.

2.4. Tipos de riesgos asociados a la Anonimización de Datos.

La anonimización de datos implica considerar varios tipos de riesgos que pueden surgir al tratar de proteger la identidad de las personas en conjuntos de datos. Estos riesgos incluyen:

2.4.1. Singularización: este riesgo se refiere a la posibilidad de que los datos dentro de un conjunto puedan ser usados para identificar específicamente a una persona, basándose en información única o una combinación de atributos.

2.4.2. Vinculabilidad: Consiste en la capacidad de conectar dos o más registros que, cuando se combinan, pueden identificar a una persona específica, ya sea dentro del mismo conjunto de datos o entre diferentes.

2.4.3. Inferencia: Este riesgo se relaciona con la posibilidad de deducir detalles específicos sobre una persona utilizando los datos disponibles. A menudo, combinando varios atributos, se puede inferir información sensible o detallada, incluso cuando los datos han sido modificados o desidentificados.

3. Proceso de Anonimización.

El proceso de anonimización comprende varias etapas esenciales. Esta guía está diseñada para ofrecer orientación y no aborda aspectos técnicos detallados. En lugar de estar centrada en metodologías específicas para la anonimización de datos, está enfocada en describir las etapas generales del proceso. Estas etapas incluyen la identificación de los objetivos que se quieren alcanzar con los datos anonimizados, el conocimiento de los datos a tratar, seguido por la desidentificación de los mismos. Posteriormente, se revisan las distintas técnicas de anonimización y se evalúan los riesgos de identificación asociados con cada una de ellas, para finalizar con ejemplos prácticos de aplicación de estas técnicas en los casos de uso previamente definidos.

Se puede utilizar este paso a paso como orientación general para anonimizar los conjuntos de datos cuando sea apropiado, dependiendo del caso de uso específico.

3.1. Identifique los objetivos que se quieren alcanzar con los datos anonimizados.

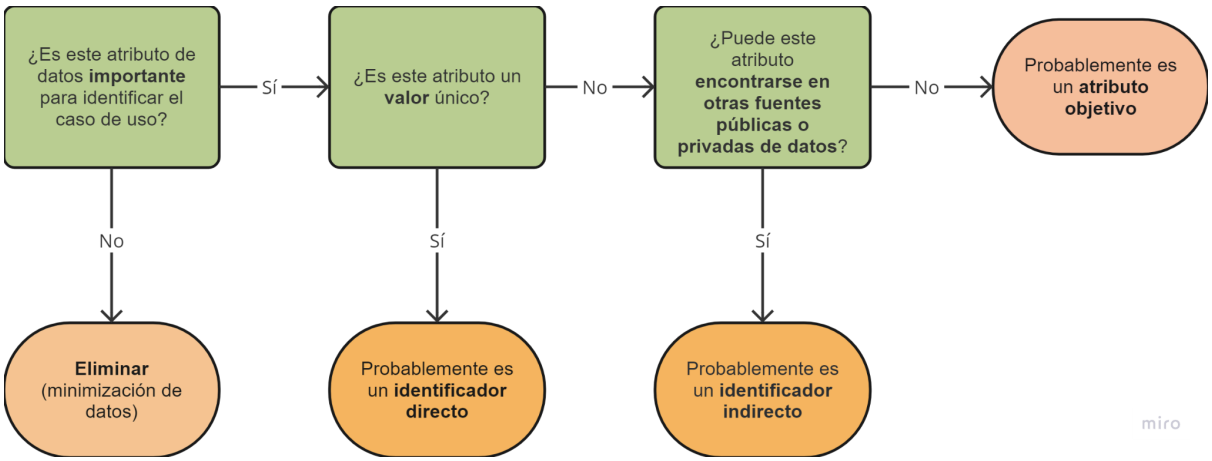
El diseño del proceso de anonimización debe estar claramente orientado hacia los objetivos específicos que se pretenden alcanzar con el uso de los datos anonimizados. Dependiendo del propósito final, como se mencionó anteriormente, los datos anonimizados pueden destinarse a ser publicados como datos abiertos o utilizados internamente bajo condiciones más restringidas. En casos donde los datos se manejen con restricciones, es crucial incorporar acuerdos de confidencialidad que incluyan cláusulas detalladas sobre la reidentificación y la protección de la privacidad de la información.

Además, es fundamental definir claramente la unidad y los funcionarios dentro de la organización que serán responsables del proceso de anonimización. Estos responsables deben tener un conocimiento profundo de los contenidos de la base de datos para asegurar un manejo adecuado y seguro de la información a lo largo de todo el proceso.

3.2. Conozca sus Datos.

Una base de datos estructurada que contiene información personal se compone de múltiples atributos con diversos grados de capacidad para identificar a un individuo. El proceso de anonimización involucra normalmente la supresión de identificadores que permiten una conexión directa con la persona y la modificación de aquellos que podrían hacerlo de manera indirecta. Sin embargo, se mantienen aquellos atributos cruciales para la operatividad del sistema, salvo que el objetivo sea la creación de datos completamente sintéticos.

A continuación, presentamos un diagrama de flujo diseñado para ayudar en la adecuada clasificación de los atributos de sus datos, un paso preliminar esencial antes de avanzar hacia la desidentificación efectiva de los mismos.



3.3. Desidentificación de los datos.

La desidentificación es fundamental en el proceso de anonimización, orientada a balancear la utilidad de los datos con las necesidades de privacidad. Este paso implica la eliminación o modificación de identificadores, tanto directos como indirectos, dependiendo del nivel de privacidad requerido y del contexto específico en el que se utilizarán los datos.

En situaciones donde los datos están destinados a ser compartidos o sujetos a análisis exhaustivos, simplemente remover identificadores directos podría no ser suficiente. En estos casos, es imperativo evaluar y ajustar también los identificadores indirectos que, al estar interrelacionados con otras fuentes de datos o combinados con otros atributos, podrían posibilitar la reidentificación de individuos. La adecuada gestión de estos identificadores es esencial para prevenir la exposición no deseada de la identidad personal.

En el siguiente ejemplo, el identificador directo “nombre” ha sido eliminado, para proteger la identidad de las personas[*].

Nombre	Edad	Serie Favorita
José	54	The Big Bang Theory
Pedro	30	Friends
María	23	Stranger things

**No obstante, existe el riesgo de que individuos como José, Pedro y María puedan ser reidentificados si sus registros desidentificados son combinados con otras fuentes de datos disponibles públicamente, tales como sus perfiles en redes sociales.*

En situaciones específicas, asignar un seudónimo a cada registro puede ser útil para mantener vínculos con el individuo original sin comprometer su identidad. Aquí hay algunos ejemplos:

- Fusión de Datos: Al combinar registros de diversas fuentes, el uso de seudónimos permite mantener asociaciones sin exponer la identidad real del individuo, facilitando análisis integrales sin riesgo de violar la privacidad.
- Análisis de Múltiples Registros: Cuando se evalúan múltiples registros relacionados con un individuo, los seudónimos ayudan a mantener la confidencialidad mientras se permite el análisis profundo de los datos.
- Generación de Datos Sintéticos: Emplear seudónimos en lugar de identificadores directos es crucial para desarrollar aplicaciones y realizar pruebas, asegurando que la privacidad se mantenga incluso en entornos de prueba.

Es fundamental que los seudónimos asignados sean únicos para cada identificador directo y que su diseño impida la reversión por terceros no autorizados. Por ejemplo, un seudónimo robusto podría ser un código generado aleatoriamente que no guarde relación matemática o lógica con el dato original, como cambiar el nombre "José" por un número único "1234". Sin embargo, para aumentar aún más la seguridad, podría utilizarse una combinación más compleja de caracteres alfanuméricos y símbolos, como "A3#m9Z!", que es aún menos susceptible a ser rastreado o deducido.

La siguiente tabla ilustra ejemplos cómo se asignan seudónimos robustos y únicos a cada identificador directo:

Nombre	Token	Edad	Serie Favorita
José	1234	54	The Big Bang Theory
Pedro	A3#m9Z!	30	Friends
María	98gv$$M	23	Stranger things

Para asegurar la posibilidad de relacionar registros desidentificados con sus correspondientes identificadores originales en el futuro, es fundamental mantener una tabla segura que vincule estos identificadores directos con sus respectivos seudónimos. Esta tabla no solo es crucial para el cumplimiento de requisitos legales y de privacidad, sino que también facilita la gestión cotidiana de la base de datos por parte de su propietario. La seguridad de esta tabla se asegura mediante su almacenamiento en sistemas cifrados y restringiendo el acceso a un grupo limitado de personal autorizado, garantizando así la integridad y confidencialidad de los datos en todo momento.

3.4. Técnicas de anonimización:

En el proceso de anonimización, se emplean una variedad de técnicas diseñadas para abordar los distintos aspectos de dicho proceso, todas con el objetivo de minimizar los riesgos de reidentificación. Existen múltiples enfoques generales, cada uno compuesto por técnicas específicas cuya aplicabilidad puede diferir considerablemente dependiendo del tipo de datos y del contexto específico en el que se implementan. Debido a que la efectividad de estas técnicas puede estar fuertemente influenciada por el entorno legislativo y las circunstancias particulares, es recomendable que cada entidad consulte con expertos en esta materia para determinar la técnica más adecuada a sus necesidades específicas.

En esta sección, se proporciona un resumen de dichas técnicas, clasificadas según su enfoque y se ofrecen orientaciones generales sobre las circunstancias más apropiadas para su empleo.

3.4.1. Enmascaramiento (*masking)*:

El enmascaramiento es una técnica de protección de la identidad que involucra ocultar o eliminar información para evitar la identificación directa de los individuos. Se utiliza principalmente en contextos donde es crítico proteger la privacidad de los datos personales.

3.4.1.1. Supresión de registros:

Descripción	La supresión de registros se refiere a la eliminación de un registro completo en un conjunto de datos. A diferencia de la mayoría de las otras técnicas, esta técnica afecta a múltiples atributos al mismo tiempo.

Cuándo usarlo	Esta técnica es útil para eliminar registros atípicos que son únicos o que no cumplen otros criterios dentro del conjunto de datos anonimizado. Los valores atípicos pueden ser más fáciles de reidentificar, ya que se destacan del resto de los datos. La supresión puede aplicarse antes o después de haber implementado otras técnicas de anonimización como la generalización.
Cómo usarlo	La recomendación es eliminar por completo el registro del conjunto de datos. Es importante recordar que la supresión debe ser permanente y no debe concebirse como simplemente ocultar temporalmente el registro. Además, reducir la visibilidad de los datos sin eliminarlos completamente puede no ser suficiente para proteger contra la reidentificación si los datos subyacentes siguen siendo accesibles.
Otros consejos	Hay que considerar que la eliminación de registros puede tener impactos en el análisis de los datos, afectando, por ejemplo, la precisión estadística como el promedio y la mediana del conjunto de datos.

3.4.1.2. Enmascaramiento de caracteres:

Descripción	Sustitución de caracteres originales por alternativos, como asteriscos o "X", para ocultar información sensible sin alterar la estructura general del dato.
Cuando usarlo	Útil para proteger información como números de identificación personal o direcciones de email, cuando solo una parte del dato es sensible y el resto puede mantenerse visible para reducir el riesgo de identificación.
Cómo usarlo	Elija los caracteres específicos dentro de un atributo que necesitan ser ocultados y reemplácelos manteniendo el formato original del dato.
Otros consejos	Asegúrese de que el método de enmascaramiento no sea predecible y que mantenga la irrelevancia de los caracteres ocultos en cualquier análisis, evitando así patrones que podrían ser descifrados fácilmente. Considerar el enmascaramiento completo en situaciones donde se desea ocultar totalmente la información sensible.

3.4.1.3. Cifrados:

Descripción	Convierte datos legibles en un formato codificado que solo se puede revertir a su estado original mediante una clave específica, protegiendo así la información durante su transmisión y almacenamiento.
Cuando usarlo	Apropiado para la transmisión de datos sensibles a través de redes

	inseguras o en sistemas vulnerables a accesos no autorizados.
Cómo usarlo	Cifre datos sensibles antes de transmitirlos o almacenarlos, utilizando algoritmos de cifrado robustos y gestionando adecuadamente las claves de cifrado para evitar su exposición o pérdida.
Otros consejos	● Revisar periódicamente la robustez de los algoritmos de cifrado y adaptarlos si es necesario frente a avances tecnológicos que puedan afectar su seguridad. ● Implementa políticas de rotación de claves para minimizar el riesgo de exposición de datos sensibles. ● Asegurar que las políticas de cifrado y descifrado estén actualizadas para evitar posibles vulnerabilidades. ● Evaluar regularmente el uso de cifrados en todos los sistemas de datos para garantizar una protección continua y efectiva.

3.4.2. Aleatorización:

Este tipo de técnicas se basan en modificar o alterar la veracidad de los datos a nivel individual, respetando la distribución global de éstos, consiguiendo reducir así la vinculabilidad y la inferencia. Sin embargo, es importante destacar que la aleatorización por sí sola no garantiza la protección contra la identificación individual. Se recomienda combinarla con otros métodos, como el filtrado de atributos o la generalización, para fortalecer la privacidad de los datos.

3.4.2.1. Adición de ruido :

Descripción	Modifica los datos originales con variaciones aleatorias para disociarlos de cualquier identificador, manteniendo su utilidad general. Aunque los valores exactos varíen, los datos permanecen útiles para análisis.
Cuando usarlo	Esta técnica es ideal cuando los datos necesitan mantener su estructura básica para análisis pero requieren una reducción en la precisión para prevenir identificación directa.
Cómo usarlo	Aplicación consistente del ruido: introducir ruido uniformemente y con precisión a lo largo de los datos para prevenir la revelación de información personal. Mantenimiento de la utilidad: ajustar el nivel de ruido para no comprometer la calidad o utilidad analítica de los datos. Evaluación rigurosa: realizar revisiones periódicas para asegurar que el nivel de ruido sigue siendo efectivo en proteger la privacidad sin comprometer la utilidad.

Otros consejos	Aplica el ruido de manera uniforme y calculada para que no revele información personal directa. Asegurar que la calidad y la utilidad de los datos no se vean comprometidas.

3.4.2.2. Permutación o intercambio :

Descripción	Mezcla valores de atributos en un conjunto de datos, asignando atributos de manera aleatoria a diferentes registros. Esto mantiene la distribución general, pero previene la vinculación directa a individuos específicos, conservando la estructura estadística sin revelar identidades.
Cuando usarlo	Útil en análisis estadísticos donde es crucial ocultar identidades, como en conjuntos de datos financieros o médicos donde las interrelaciones son críticas.
Cómo usarlo	Mantener correlaciones lógicas: Asegurar que la permutación preserve la estructura de datos y lógica interna. Selección cuidadosa de atributos: Evita permutar atributos objetivo que puedan presentar riesgos de identificación. Control Riguroso: Implementar controles para evitar errores o sesgos en los resultados.
Otros consejos	• Validación de la permutación: Realizar pruebas para asegurarse de que no se introducen distorsiones. • Combinar con otras técnicas: Usar permutación con otras técnicas de anonimización para una protección más robusta. • Revisión periódica: Revisar periódicamente las estrategias de permutación para asegurar su efectividad.

3.4.2.3. Privacidad diferencial:

Descripción	Privacidad diferencial es una técnica que añade ruido deliberadamente a los datos, asegurando que los resultados sean menos identificables a nivel individual mientras protege la información personal.
Cuando usarlo	Ideal para entornos que requieren compartir datos agregados sin revelar identidades, esta técnica asegura que los resultados de análisis protejan la privacidad individual sin sacrificar la utilidad de los datos.
Cómo usarlo	Determinar la cantidad de ruido necesario: establecer la cantidad de ruido necesaria para proteger la privacidad sin comprometer la integridad de los datos.

	Supervisión continua: ajustar el nivel de ruido según sea necesario para asegurar que los datos modificados preserven la calidad y precisión esperada.
Otros consejos	Seguimiento riguroso: mantener un registro detallado de las consultas para garantizar la adecuación de las respuestas y la efectividad de la protección. Evitar motores de búsqueda abiertos: no utilizar motores de búsqueda que faciliten trazabilidad y puedan comprometer la privacidad. Generar resultados estadísticos: Limitar los resultados a datos generales para minimizar el riesgo de identificación.

3.4.3. Generalización :

La generalización es una técnica de anonimización que diluye los atributos específicos de los datos, ajustando sus escalas o magnitudes. Por ejemplo, se puede reemplazar el nombre de una ciudad con el de una región más amplia, o cambiar una fecha específica por el mes correspondiente. Aunque este método puede ayudar a evitar la identificación directa de individuos, no garantiza por sí solo una anonimización completa y efectiva. Es necesario aplicar el proceso de forma adecuada y aplicar otras técnicas de forma conjunta para garantizar la protección de los datos personales.

3.4.3.1. K- Anonimidad :

Descripción	El modelo de k-anonimato diluye atributos específicos para generalizar los datos, asegurando que todos los registros dentro de una misma categoría sean indistinguibles entre sí por contener al menos k registros similares. Este parámetro k, conocido como el factor de anonimato, es crucial para proteger la privacidad, ya que evita que los registros puedan ser directamente vinculados a individuos específicos y protege contra riesgos de vinculación y reidentificación.
Cuando usarlo	Útil para compartir datos dentro de un grupo sin revelar identidades, especialmente en grandes volúmenes de datos donde la privacidad es crucial.
Cómo usarlo	Establecer el valor de k: seleccionar un valor de k que ofrezca un equilibrio entre proteger la privacidad y mantener la utilidad de los datos, asegurando que cada grupo o clase de equivalencia contenga al menos k registros similares. Implementación: Ajustar los registros de manera que cada clase de equivalencia mantenga uniformidad, haciendo que los registros dentro de cada grupo sean indistinguibles.

Otros consejos	Monitoreo de efectividad: es fundamental revisar la eficacia de la técnica regularmente para adaptarse a cambios en los volúmenes de datos y en las necesidades de análisis. Validación post-anonimización: realizar evaluaciones periódicas para asegurar que las modificaciones no introduzcan distorsiones que comprometan la integridad de los datos. Consideraciones de equilibrio: tener en cuenta que un k muy alto puede reducir la utilidad práctica de los datos, mientras que un k muy bajo podría comprometer la privacidad.

3.4.3.2. Diversidad-L y Proximidad-T:

Descripción	Las técnicas de Diversidad-L y Proximidad-T son extensiones del modelo de K-anonimato, diseñadas para mejorar la protección contra ataques de inferencia y vinculación. Diversidad-L asegura que cada atributo dentro de un grupo tenga valores únicos, mejorando la distribución de datos y reduciendo la identificabilidad. Proximidad-T ajusta la distribución de atributos dentro de cada grupo para mantener una distribución equilibrada, complicando cualquier intento de inferencia.
Cuando usarlo	Son útiles en bases de datos donde la preservación de los atributos estadísticos es importante sin sacrificar la protección de la privacidad.
Cómo usarlo	Configuración: elegir cuidadosamente los valores de L y T para cada conjunto de datos para mantener tanto la diversidad como la proximidad necesarias. Verificación y ajuste: asegurar que las clases de equivalencia cumplan con los estándares de anonimato y ajuste según sea necesario para mantener la efectividad.
Otros consejos	Mantenimiento de la distribución uniforme: asegurar que los valores dentro de cada grupo sean uniformes para evitar patrones explotables. Revisión periódica: es crucial revisar y actualizar periódicamente la implementación para asegurar que se sigue proporcionando el nivel de privacidad deseado frente a las nuevas técnicas de análisis de datos.

3.4.4. Seudonimización :

Descripción	Implica sustituir identificadores de datos por valores alterados, como los generados mediante algoritmos de cifrado o hash. Aunque este proceso no elimina totalmente la posibilidad de identificación, reduce

	significativamente la facilidad con la que se pueden vincular los datos alterados a una persona específica. Este método mejora la seguridad de los datos, manteniendo su utilidad para el análisis sin comprometer la identidad original.
Cuando usarlo	Es efectivo en escenarios donde es crucial mantener cierto grado de vinculación entre datos para el procesamiento a través de sistemas, sin revelar la identidad. Se utiliza ampliamente en ambientes donde la privacidad y el acceso seguro a datos para análisis son críticos.
Cómo usarlo	● Cifrado con clave secreta: aplica cifrado a los datos utilizando una clave conocida sólo por el responsable de tratamiento, protegiendo los datos durante su almacenamiento y transmisión. ● Función hash: convierte datos a un valor numérico fijo, haciendo difícil su reconversión a la forma original sin una clave o contexto específico. ● Función hash con clave almacenada: similar al hash regular, pero incluye una clave secreta que mejora la seguridad del proceso de hash. ● Descomposición en tokens: sustituye identificadores directos por tokens que no se relacionan directamente con los datos originales pero permiten análisis útiles sin revelar información sensible.
Otros consejos	Monitoreo y ajuste continuo: verifica regularmente la eficacia de las técnicas de seudonimización para contrarrestar amenazas emergentes y adaptarse a los cambios tecnológicos. Diversificación de claves y seudónimos: emplea diferentes claves y seudónimos en distintos conjuntos de datos para minimizar riesgos de vinculación y mejorar la seguridad. Protección de claves: asegura que las claves utilizadas en el proceso de seudonimización eviten el acceso no autorizado, especialmente si se utilizan claves reversibles.

3.5. Riesgo de Identificación de las técnicas de anonimización:

Una vez comprendidas las técnicas de anonimización, es importante reconocer y evaluar los riesgos de identificación que podrían surgir de la aplicación de dichas técnicas. El riesgo de anonimización es la probabilidad de reidentificar a un individuo u organización dentro de un conjunto de datos. Cada método de anonimización tiene fortalezas y debilidades en cuanto a la protección de la identidad, variando en su capacidad para minimizar los riesgos de singularización, vinculabilidad e inferencia. Comprender estos riesgos es esencial para elegir la técnica más adecuada según elcontexto y los requerimientos específicos de privacidad y uso de datos.

La siguiente tabla resume cómo diferentes técnicas de anonimización abordan estos riesgos, ofreciendo una comparativa que facilita la elección informada de la técnica más apropiada según el contexto específico de uso.

Tabla de Riesgos de Identificación por Técnica de Anonimización

Tipo de técnica	Técnica	Riesgo de Singularización	Riesgo de Vinculabilidad	Riesgo de Inferencia
Enmascaramiento	Supresión de registros	Sí	Sí	Sí
	Enmascaramiento de caracteres	Sí	Sí	Sí
	Cifrados	Sí	Sí	A veces
Aleatorización	Adición de ruido	Sí	A veces	A veces
	Permutación o intercambio	Sí	A veces	A veces
	Privacidad diferencial	A veces	A veces	A veces
Generalización	Anonimato-K	No	Sí	Sí
	Diversidad l	No	Sí	A veces
	proximidad t	No	Sí	A veces
Seudonimización	Cifrado por hash	Sí	Sí	A veces
	Tokenización	Sí	Sí	Sí

Explicación de los Niveles de Riesgo:

Sí: la técnica no mitiga eficazmente el riesgo y requiere el uso combinado con otras técnicas para mejorar la protección.
A veces: la técnica puede mitigar el riesgo en ciertas circunstancias o hasta cierto grado. Es importante considerar el contexto específico en el que se aplicará la técnica.
No: la técnica es efectiva para mitigar el riesgo asociado, ofreciendo una protección robusta contra el tipo de riesgo especificado.

Aunque no es el foco principal de esta guía, es altamente recomendable calcular el riesgo de identificación antes de aplicar técnicas de anonimización propiamente dichas, utilizando alguna metodología de medición del riesgo. Esta práctica previa permite evaluar y entender el potencial de reidentificación asociado a los datos en su forma original y cómo este riesgo podría modificarse mediante diversas técnicas de anonimización. Implementar una evaluación rigurosa del riesgo no solo ayuda a seleccionar la técnica más adecuada para cada conjunto de datos, sino que también guía la configuración de los parámetros de anonimización para maximizar la efectividad de la protección de la privacidad sin comprometer la utilidad de los datos. Además, realizar una evaluación continua del riesgo durante y después del proceso de anonimización asegura que los datos permanezcan seguros frente a nuevas vulnerabilidades y técnicas de análisis que puedan surgir.

3.5.1. Aplicación técnicas de anonimización:

Durante esta etapa final del proceso de anonimización, se seleccionan y aplican técnicas específicas para tratar tanto los identificadores indirectos como los atributos objetivos. El objetivo es impedir que estos se asocien fácilmente con datos personales identificables. Ajustar cuidadosamente los valores de los datos es esencial para mantener su relevancia y utilidad en análisis específicos.

Es vital optar por técnicas que logren un equilibrio entre la protección de la privacidad y las capacidades analíticas de los datos. Según las necesidades y características de cada caso, se podrán implementar estrategias de anonimización variadas.

Ejemplos de Casos de Uso con Técnicas de Anonimización Recomendadas:

Casos de uso	Técnicas de anonimización sugeridas para datos a nivel de registro

Publicación de Datos Abiertos	Supresión de registros, Anonimato K, Diversidad L, proximidad T
Compartir datos entre Organizaciones	Supresión de registros, Anonimato K, Diversidad L, proximidad T
Intercambio interno de datos (datos desidentificados)	Supresión de registros, Enmascaramiento, Cifrado, Seudonimización
Intercambio interno de datos (datos anonimizados)	Supresión de registros, Enmascaramiento de caracteres, Adición de ruido, Permutación o intercambio, Privacidad Diferencial

Cabe destacar que los ejemplos proporcionados son ilustrativos, según las necesidades y el contexto, es posible optar por otras técnicas de anonimización.

4. Finalización y Mantenimiento del Proceso de Anonimización

Una vez aplicadas las técnicas de anonimización, es importante llevar a cabo una evaluación rigurosa de la utilidad de los datos anonimizados. Esta evaluación debe equilibrar la necesidad de proteger la privacidad con la capacidad analítica de los datos, asegurando que los datos sigan siendo útiles para los propósitos previstos sin comprometer la identidad de los individuos. Documentar meticulosamente este balance, junto con las técnicas aplicadas y cualquier ajuste realizado, es esencial para la auditoría y replicabilidad del proceso.

Como se mencionó anteriormente, aunque esta guía no propone una metodología específica para la evaluación del riesgo de identificación debido a que está fuera de su alcance, es crucial que cada entidad defina y aplique su propio método para evaluar y monitorear constantemente este riesgo. Este enfoque asegura que la anonimización se mantenga efectiva ante cambios en los contextos de uso de los datos o en las tecnologías de análisis, proporcionando así una protección continua de la privacidad a lo largo del tiempo.

5. Herramientas

Para facilitar y mejorar el proceso de anonimización, a continuación se presentan dos herramientas de software libre diseñadas para facilitar la anonimización de datos personales de manera eficaz y sin costo.

Herramienta	Descripción	URL
ARX	ARX es un software de código abierto para anonimizar datos personales.	https://arx.deidentifier.org/
Amnesia	La herramienta de anonimización Amnesia es un software utilizado localmente para anonimizar datos personales. Actualmente admite garantías de k-anonimidad y km-anonimidad.	https://amnesia.openaire.eu/

6. Conclusión.

La presente guía proporciona una visión general sobre las técnicas de anonimización de datos, con el objetivo de ofrecer a las entidades y organizaciones una introducción básica y práctica para la gestión segura de datos personales. A lo largo de este documento, se han explorado diversos enfoques y algunas herramientas que pueden facilitar la implementación de estrategias de anonimización efectivas y eficientes, enfocándonos en opciones accesibles y sin costo..

Es importante destacar que, aunque la guía ofrece una introducción sólida sobre el proceso de anonimización, no propone una metodología específica para la evaluación del riesgo de identificación ni detalla en profundidad el cómo aplicar las técnicas, dado que esto podría exceder el alcance previsto. En su lugar, se recomienda encarecidamente que cada organización trabaje en colaboración con expertos en anonimización y protección de datos para adaptar las técnicas discutidas a sus necesidades específicas y contextos operativos, garantizando así que las prácticas de anonimización se mantengan efectivas y actualizadas frente a los cambios tecnológicos y normativos.

Por último, esta guía debe considerarse un punto de partida para la educación y la toma de conciencia sobre la importancia de la protección de la privacidad en la era digital. El compromiso continuo y la adaptación a los nuevos desafíos en la gestión de datos personales son esenciales para asegurar la efectividad de cualquier estrategia de anonimización, cumpliendo con las regulaciones vigentes y respetando el derecho a la privacidad de los individuos, y así mantenerse efectiva ante los desafíos cambiantes de la gestión de datos en el mundo digital.

7. Referencias.

- Autoridad Nacional de Protección de Datos de Singapur. (2022). Guía básica de anonimización. PDPC - Personal Data Protection Commission Singapore. Disponible en: https://www.aepd.es/documento/guia-basica-anonimizacion.pdf
- Agencia Española de Protección de Datos. (2016). Orientaciones y garantías en los procedimientos de anonimización de datos personales. Disponible en: https://www.aepd.es/guias/guia-orientaciones-procedimientos-anonimizacion.pdf
- Montaña, C., & Rodríguez, B. (2017). Criterios de disociación de datos personales (2ª ed.). Agencia para el Desarrollo del Gobierno de Gestión Electrónica y la Sociedad de la Información y el Conocimiento (AGESIC). Disponible en: https://www.gub.uy/agencia-gobierno-electronico-sociedad-informacion-conocimiento/sites/agencia-gobierno-electronico-sociedad-informacion-conocimiento/files/documentos/noticias/11--criterios-de-disociacion-de-datos-personales.pdf
- Personal Data Protection Commission. (2024). Advisory guidelines on the PDPA for selected topics (Revised 17 May 2024). Disponible en: https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/advisory-guidelines/ag-on-selected-topics/advisory-guidelines-on-the-pdpa-for-selected-topics-(revised-may-2024).pdf
- Elliot, M., Mackey, E., & O'Hara, K. (2020). The Anonymisation Decision-Making Framework: European Practitioners' Guide (2nd ed.). Manchester, UK: UKAN Publications, University of Manchester. Disponible en: https://eprints.soton.ac.uk/445373/1/adf_2nd_edition_1.pdf
- Ministerio de Asuntos Económicos y Transformación Digital. Gobierno de España (2022). Introducción a la anonimización: Técnicas y casos prácticos. Disponible en: https://datos.gob.es/sites/default/files/doc/file/introduccion_a_la_anonimizacion_de_datos-_tecnicas_y_casos_practicos_1.pdf
- Instituto Nacional de Estadísticas. Chile (2021). Guía para el control de divulgación estadística en microdatos. Disponible en: + https://www.ine.gob.cl/docs/default-source/buenas-practicas/estandares/estandar/documento/guía-control-divulgación-estadística-microdatos.pdf?sfvrsn=fb568638_2
- Gil González, E. (2016). Big data, privacidad y protección de datos. Madrid: Agencia Española de Protección de Datos y Agencia Estatal Boletín Oficial del Estado. Disponible en: https://www.aepd.es/media/premios/big-data.pdf
- Comisión Europea Grupo de trabajo sobre protección de datos del artículo 29. (2014). Dictamen 05/2014 sobre técnicas de anonimización. Bélgica. Disponible en: https://www.aepd.es/documento/wp216-es.pdf

- Ministerio Secretaría General de la Presidencia. (1999). Ley sobre protección de la vida privada (Ley N°. 19628). Biblioteca del Congreso Nacional de Chile. Disponible en: https://www.bcn.cl/leychile/navegar?idNorma=141599&idVersion=2020-08-26