agencia española protección datos

Guía básica de anonimización

Elaborada por Autoridad Nacional de Protección de Datos de Singapur

(PDPC - Personal Data Protection Commission Singapore)

Guía publicada en
octubre 2022

PRÓLOGO AEPD

El presente documento es la traducción de la guía básica elaborada por la Autoridad Nacional de Protección de Datos de Singapur (PDPC), que se puede descargar en su versión original en ingles en los siguientes enlaces.

PDPC | Basic Anonymisation
Guide-to Basic-Anonymisation-31-March-2022.ashx (pdpc.gov.sg)

La División de Innovación Tecnológica de la AEPD estimó el gran valor didáctico de esta guía por lo que, tras pedir las autorizaciones correspondientes al PDPC Singapore, se decidió a la traducción y publicación en su espacio de divulgación dentro de la página web de la AEPD.

Lo mismo sucedió con la herramienta de anonimización, también elaborada por la Autoridad Nacional de Protección de Datos de Singapur (PDPC) y que se puede descargar del primer enlace.

Este documento junto con su herramienta de anonimización se considera de especial interés para responsables, encargados de tratamientos y delegados de protección de datos (DPD).

Palabras clave: protección de datos, responsabilidad proactiva, riesgo, anonimización, seudonimización, protección de datos desde el diseño, DPD,
seguridad.

ÍNDICE

1. INTRODUCCIÓN DE LA GUÍA

Esta guía está destinada a proporcionar una introducción y orientación práctica a las organizaciones que son nuevas en la anonimización sobre cómo realizar adecuadamente la anonimización básica y la desidentificación de conjuntos de datos estructurados[1] , textuales[2] y no complejos[3] . Presenta el flujo de trabajo de anonimización en el contexto de cuatro casos de uso comunes.

Esta guía no es exhaustiva al tratar todas las cuestiones relacionadas con la anonimización, la desidentificación y la reidentificación de los conjuntos de datos. Se aconseja a las organizaciones que consideren la contratación de expertos en anonimización, estadística o evaluadores de riesgos independientes para realizar las técnicas de anonimización adecuadas o la evaluación de los riesgos de reidentificación, cuando los problemas de anonimización sean complejos (por ejemplo, grandes conjuntos de datos que contengan una amplia gama de datos personales longitudinales o pertenecientes a categorías especiales).

La implementación de las recomendaciones de esta guía no implica el cumplimiento de la normativa de protección de datos.

Las diferentes jurisdicciones ven la anonimización de manera diferente y, por lo tanto, las recomendaciones proporcionadas en esta guía pueden no aplicarse a las leyes de protección de datos en otros países.

Esta guía debe leerse junto con el documento del PDPC: **“las Directrices de asesoramiento del PDPC sobre la Ley de Protección de Datos Personales para Temas Seleccionados”.**

1 Estructurado” se refiere a los datos en un formato definido y tabular, como una hoja de cálculo o una base de datos relacional (por ejemplo.XLSX y CSV).

2 “Textual” se refiere a texto, números, fechas, etc., es decir, datos alfanuméricos ya en forma digital. Las técnicas de anonimización para datos no textuales como audio, video, imágenes, datos biométricos, etc., crean desafíos adicionales y requieren diferentes técnicas de anonimización, que están fuera del alcance de esta guía.

3 El enfoque recomendado en esta guía se aplica solo al tratamiento de datos estructurados (es decir, datos textuales en forma tabular en columnas y filas, como hojas de cálculo de Microsoft Excel). Las fotografías digitales, por ejemplo, no entran en esta categoría de datos.

2. ANONIMIZACIÓN VERSUS DESIDENTIFICACIÓN

La **anonimización** consiste en la conversión de datos personales en datos que no se pueden utilizar para identificar a ningún individuo. La anonimización hay que considerarla como un proceso basado en el riesgo, que incluye tanto la aplicación de técnicas de anonimización como salvaguardas para evitar la reidentificación.

La **desidentificación**[4] consiste en la eliminación de identificadores (por ejemplo, nombre, dirección, número de documento nacional de identidad) que identifican directamente a un individuo. La desidentificación a veces se entiende erróneamente con la anonimización, sin embargo, es solo el primer paso de la anonimización. Un conjunto de datos desidentificado puede volver a identificarse fácilmente cuando se combina con datos que son de acceso público o fácil.

La **reidentificación** se refiere a la identificación de individuos a partir de un conjunto de datos que previamente fue desidentificado o anonimizado.

Los datos anonimizados no se consideran datos personales y, por lo tanto, no se rigen por la normativa de protección de datos.

1. UN EJEMPLO DE DESDENTIFICACIÓN

Albert usa aplicaciones de pedidos de comida a domicilio con frecuencia. Su aplicación favorita de pedidos de comida, SuperHungry, decide publicar información sobre sus usuarios para un hackathon[5].

Registro de datos de Albert en SuperHungry

Nombre	Restaurante favorito	Comida favorita
Alberto Ruiz	Restaurante Katong	Combo de 3 piezas de pollo

Fecha de nacimiento	Género	Empresa
01/01/1990	Hombre	ABC Pte Ltd

4 El proceso de desidentificación puede incluir la asignación de seudónimos.

5 Un **hackathon** (o **hackatón**) consiste en un encuentro de programadores cuyo objetivo es el desarrollo colaborativo de software, aunque en ocasiones puede haber también un componente de hardware. El objetivo es doble: por un lado, hacer aportes al proyecto de software libre que se desee y, por otro, aprender sin prisas.

SuperHungry desidentifica el conjunto de datos eliminando los nombres antes de publicarlo, pensando que esto equivale a anonimizar el conjunto de datos.

El registro desidentificando de Albert publicado por SuperHungry

Nombre	Restaurante favorito	Comida favorita	Fecha de nacimiento	Género	Empresa
Alberto Ruiz	Restaurante Katong	Combo de 3 piezas de pollo	01/01/1990	Hombre	ABC Pte Ltd

Sin embargo, Albert puede ser reidentificado combinando su registro desidentificado con otros registros (por ejemplo, información personal de su perfil de redes sociales).

Perfil de Albert en las redes sociales:

Nombre	Fecha de nacimiento	Género	Empresa
Alberto Ruiz	01/01/1990	Hombre	ABC Pte Ltd

Cualquier persona con suficiente motivación puede identificar[6] fácilmente a una persona como Albert a partir de los datos desidentificados si hay otra información pública o fácilmente disponible para permitir dicha reidentificación. Si el conjunto de datos o el conjunto de datos combinado es de naturaleza sensible, se requerirá una mayor anonimización.

6 En este ejemplo se vuelve a identificar el registro sólo si esta información es exclusiva de Albert en la población.

3. INTRODUCCIÓN A LOS CONCEPTOS BÁSICOS DE ANONIMIZACIÓN DE DATOS

La anonimización de datos requiere una buena comprensión de los siguientes elementos, que deben tenerse en cuenta al determinar cuáles son las técnicas de anonimización adecuadas y los niveles de anonimización adecuados.

A. Propósito de la anonimización y utilidad

El propósito de la anonimización debe ser claro, porque la anonimización debe hacerse específicamente para el propósito en cuestión. El proceso de anonimización, independientemente de las técnicas utilizadas, reduce en cierta medida la información original en el conjunto de datos. Por lo tanto, a medida que aumenta el grado de anonimización, la utilidad (por ejemplo, claridad y/o precisión) del conjunto de datos generalmente se reduce. Por lo tanto, la organización debe decidir el grado de compromiso entre la utilidad aceptable (o esperada) y el riesgo de reidentificación.

Cabe señalar que la utilidad no debe evaluarse a nivel de todo el conjunto de datos, ya que suele ser diferente para diferentes atributos. Un extremo sería que la precisión de un atributo de datos específico es crucial y no se debe aplicar ninguna técnica de generalización o anonimización (por ejemplo, las afecciones médicas y los medicamentos administrados a individuos pueden ser datos cruciales al analizar las tendencias de ingreso hospitalario). El otro extremo sería que el atributo de datos no sirve de nada para el propósito previsto y puede eliminarse por completo sin afectar la utilidad de los datos para el destinatario (por ejemplo, la fecha de nacimiento de las personas puede no ser importante al analizar las tendencias de las transacciones de compra).

Otra consideración importante para determinar la compensación entre utilidad y anonimización es si representa un riesgo adicional si el destinatario sabe qué técnicas de anonimización y qué grado de granularidad se han aplicado; por un lado, conocer esta información puede ayudar al analista a comprender los resultados e interpretarlos mejor, pero por otro lado puede contener pistas, lo que podría conducir a un mayor riesgo de reidentificación.

B. Reversibilidad

Por lo general, el proceso de anonimización de datos sería “irreversible” y el destinatario del conjunto de datos anonimizado no podría recrear los datos originales. Sin embargo, puede haber casos en los que la organización que aplica la anonimización conserve la capacidad de recrear el conjunto de datos original a partir de los datos anonimizados; en tales casos, el proceso de anonimización es “reversible”.

C. Características de las técnicas de anonimización

Las diferentes características de las diversas técnicas de anonimización significan que ciertas técnicas pueden ser más adecuadas para una situación o tipo de datos particular que otras. Por ejemplo, ciertas técnicas (por ejemplo, el enmascaramiento de caracteres) pueden ser más adecuadas para su uso en identificadores directos y otras (por ejemplo, agregación) para identificadores indirectos. Otra característica a considerar es si el valor del atributo es un valor continuo (por ejemplo, altura = 1,61 m) o un valor discreto (por ejemplo, “sí” o “no”), porque técnicas como la perturbación de datos funcionan mucho mejor para valores continuos.

Las diversas técnicas de anonimización también modifican los datos de maneras significativamente diferentes. Algunos modifican solo una parte de un atributo (por ejemplo, enmascaramiento de caracteres); algunos reemplazan el valor de un atributo en varios registros (por ejemplo, agregación); algunos sustituyen el valor de un atributo por un valor único, pero no relacionado (por ejemplo, seudonimización); y algunos eliminan el atributo por completo (por ejemplo, supresión de atributos).

Algunas técnicas de anonimización se pueden utilizar en combinación (por ejemplo, suprimir o eliminar registros (atípicos) después de realizar la generalización).

D. Información inferida

Es posible que cierta información se infiera de datos anónimos. Por ejemplo, el enmascaramiento puede ocultar datos personales, pero no oculta la longitud del valor original en términos del número de caracteres.

Las organizaciones también pueden considerar el orden en que se presentan los datos anonimizados. Por ejemplo, si el destinatario sabe que los registros de datos se recopilaron en orden de serie (por ejemplo, el registro de visitantes a medida que vienen), puede ser prudente (siempre que no afecte a la utilidad) reorganizar todo el conjunto de datos para evitar inferencias basadas en el orden de los registros de datos.

La inferencia no se limita a un solo atributo, sino que también puede aplicarse a todos los atributos, incluso si se hubieran aplicado técnicas de anonimización a todos. Por lo tanto, el proceso de anonimización debe tomar nota de todas las posibilidades de que se produzca una inferencia, tanto antes de decidir sobre las técnicas reales como después de aplicarlas.

E. Experiencia con el tema

Las técnicas de anonimización básicamente reducen la identificabilidad de uno o más individuos del conjunto de datos original a un nivel aceptable según las tolerancias de riesgos de la organización.

Se debe realizar una evaluación de identificabilidad y reidentificabilidad[7] antes y después de aplicar las técnicas de anonimización. Esto requiere una buena comprensión del tema al que pertenecen los datos. Por ejemplo, si el conjunto de datos son datos de atención médica, la organización probablemente requeriría que alguien con suficiente conocimiento de atención médica evalúe la singularidad de un registro (es decir, en qué grado es identificable o reidentificable).

7 “Identificabilidad” se refiere al grado en que un individuo puede ser identificado a partir de uno o más conjuntos de datos que contienen identificadores directos e indirectos, mientras que “reidentificabilidad” se refiere al grado en que un individuo puede ser reidentificado a partir de conjuntos de datos anónimos.

La evaluación antes del proceso de anonimización garantiza que la estructura y la información dentro de un atributo se identifiquen y comprendan claramente, y se evalúe el riesgo de inferencia explícita e implícita de dichos datos. Por ejemplo, un atributo que contiene el año de nacimiento proporciona implícitamente la edad, al igual que un número de documento de identidad hasta cierto punto. La evaluación después del proceso de anonimización determinará el riesgo residual de reidentificación a partir de los datos anonimizados.

Otro caso es cuando los atributos de datos se intercambian entre registros y se necesita un experto en la materia para reconocer si los registros anónimos tienen sentido.

La elección correcta de las técnicas de anonimización, por lo tanto, depende del conocimiento de la información explícita e implícita contenida en el conjunto de datos y de la cantidad o tipo de información que se pretende anonimizar.

F. Comptencia en el proceso y las técnicas de anonimización

Las organizaciones que deseen compartir conjuntos de datos anonimizados deben asegurarse de que el proceso de anonimización sea llevado a cabo por empleados que hayan recibido formación y estén familiarizados con las técnicas y principios de anonimización. Si no se encuentra la experiencia necesaria dentro de la organización, se debe contratar ayuda externa.

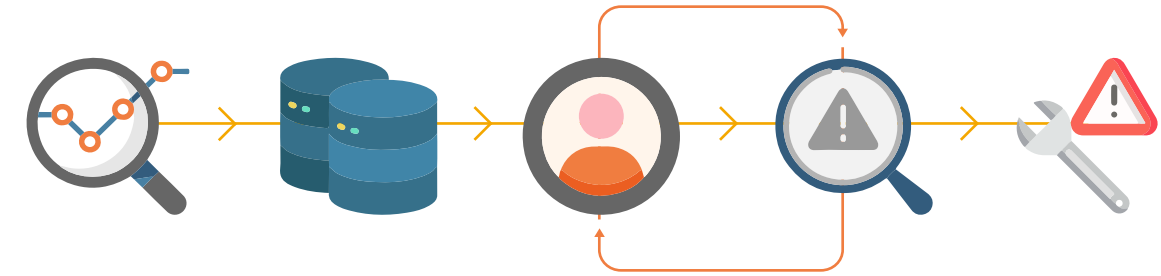
G. El destinatario

Factores como la experiencia de los destinatarios en la materia y los controles implementados para limitar la cantidad de destinatarios y evitar que los datos se compartan con partes no autorizadas desempeñan un papel importante en la elección de las técnicas de anonimización. En particular, el uso esperado de los datos anonimizados por parte del destinatario puede imponer limitaciones a las técnicas aplicadas, ya que la utilidad de los datos puede perderse más allá de los límites aceptables. Se debe tener precaución adicional al hacer públicas las divulgaciones de datos y las organizaciones requerirán un proceso más severo de anonimización en comparación con los datos compartidos en virtud de un acuerdo contractual.

H. Herramientas

Las herramientas de software pueden ser muy útiles para ayudar a ejecutar técnicas de anonimización. Consulte el anexo E para conocer algunas herramientas de anonimización disponibles en el mercado.

4. EL PROCESO DE ANONIMIZACIÓN



Paso 1	Paso 2	Paso 3	Paso 4	Paso 5
Conoce tus datos	Desidentifique sus datos	Aplique técnicas de anonimización	Calcule su riesgo	Gestione sus riesgos

Puede utilizar estos cinco pasos para anonimizar sus conjuntos de datos cuando sea apropiado, dependiendo de su caso de uso. En esta guía, explicamos estos pasos utilizando cinco casos de uso de datos que se dan comúnmente en las organizaciones.

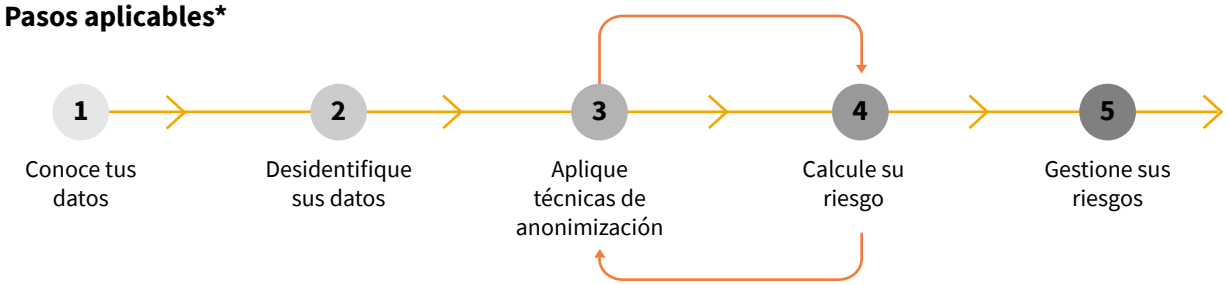
En todos los casos de uso de datos, debe asegurarse de:

- Minimización de datos, de modo que solo los atributos de datos necesarios y un extracto (cuando sea posible) de su conjunto de datos se compartan con terceros;
- Cualquier información de identificación del conjunto de datos que esté anonimizando no debe estar disponible públicamente (por ejemplo, si está anonimizando información en una base de datos de membresía, los perfiles de su base de membresía no deben estar disponibles públicamente); y
- Se otorga el nivel adecuado de protección y salvaguardia al conjunto de datos anonimizado y a la tabla de asignación de identidad para seudónimos a fin de evitar la reidentificación. En general, cuanto menos modifique un conjunto de datos a través de la anonimización, más necesitará proteger y salvaguardar el conjunto de datos, ya que los riesgos de reidentificación son mayores.

Casos de uso: Cómo se puede utilizar datos anónimos o desidentificados

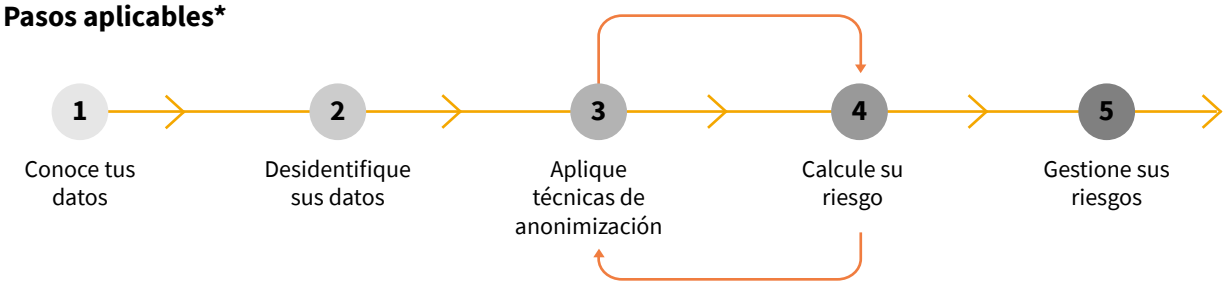
Aquí hay algunas formas en que los datos anónimos o desidentificados se pueden utilizar en su organización.

Pasos aplicables*



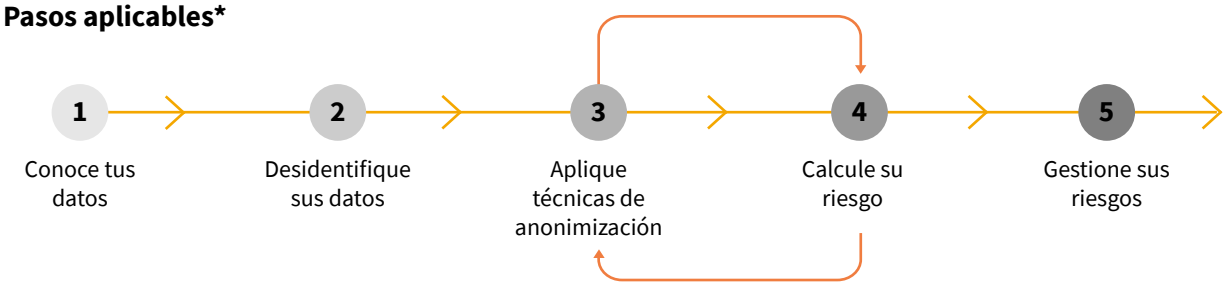
Caso de uso A	
Caso de uso	**Compartición interna de datos (datos desidentificados)** por ejemplo, datos de clientes desidentificados compartidos entre los departamentos de ventas y marketing para el análisis y el desarrollo interno de campañas de marketing dirigidas).
Descripción	Los datos solo se desidentifican para admitir el intercambio y uso de datos a nivel de registro dentro de la organización, lo que puede requerir que la mayoría de los detalles de los datos se dejen intactos. Los datos desidentificados siguen siendo datos personales, ya que es probable que sean fácilmente reidentificables. Sin embargo, sigue siendo una buena práctica desidentificar los datos, ya que proporcionan una capa adicional de protección.
¿Se necesitan controles adicionales para evitar la reidentificación?	Sí
¿El resultado final se considera datos anonimizados?	No
***Pasos aplicables**	1, 2, 5

Pasos aplicables*



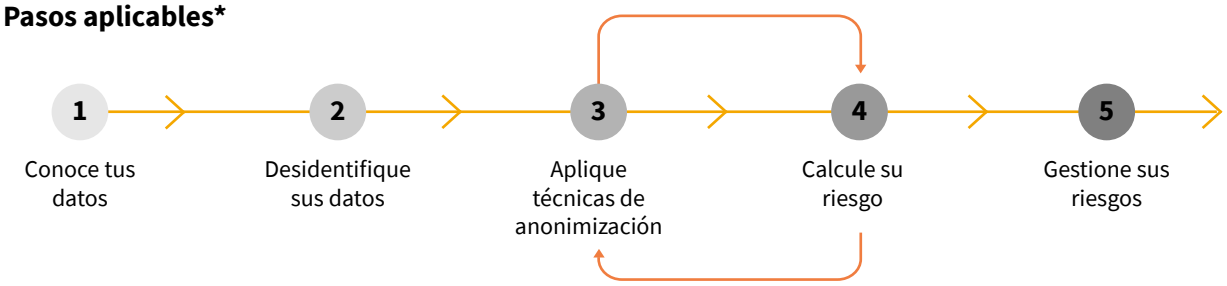
Caso de uso B	
Caso de uso	**Intercambio interno de datos (datos anonimizados)** (por ejemplo, datos anónimos sobre la demografía de los consumidores de alto valor y sus respectivos patrones de gasto compartidos con los equipos de fidelización para desarrollar propuestas de valor diferenciadas para el cliente).
Descripción	Las organizaciones podrían considerar datos anonimizados en lugar de datos desidentificados para su intercambio interno en los siguientes casos en los que: • El intercambio interno de datos no requiere datos personales detallados desidentificados (por ejemplo, para el análisis de tendencias). • Los datos involucrados son de naturaleza más sensible (por ejemplo, información financiera). • Conjuntos de datos más grandes compartidos con más de un departamento. En tales casos, las organizaciones pueden aplicar el proceso de anonimización sugerido para el intercambio de datos externos a su caso de uso interno de intercambio de datos para reducir el riesgo de reidentificación y revelación.
¿Se necesitan controles adicionales para evitar la reidentificación?	Sí
¿El resultado final se considera datos anonimizados?	Sí
***Pasos aplicables**	1, 2, 3, 4, 5

Pasos aplicables*



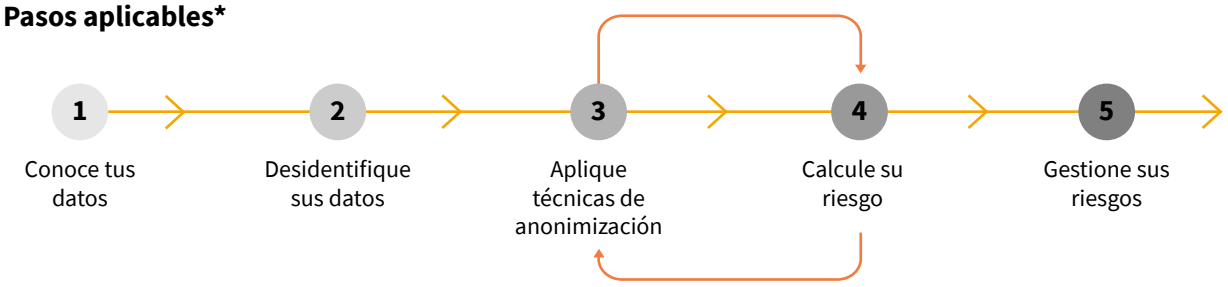
Caso de uso C	
Caso de uso	**Uso compartido de datos externos** (por ejemplo, datos anónimos de clientes compartidos entre el departamento de ventas y el socio comercial externo para el análisis de los perfiles de los clientes y el desarrollo de productos de marca compartida).
Descripción	Datos a nivel de registro compartidos con una parte externa autorizada con fines de colaboración empresarial. Las técnicas de anonimización se utilizan para convertir datos personales en datos no identificativos.
¿Se necesitan controles adicionales para evitar la reidentificación?	Sí
¿El resultado final se considera datos anonimizados?	Sí
*Pasos aplicables	1, 2, 3, 4, 5

Pasos aplicables*



Caso de uso D	
Caso de uso	**Retención de datos a largo plazo para el análisis de datos** (por ejemplo, análisis histórico de las tendencias de los clientes).
Descripción	Las técnicas de anonimización se utilizan para convertir los datos personales en datos no identificativos y permiten que los datos se mantengan a nivel de registro más allá del período de retención para el análisis de datos a largo plazo.
¿Se necesitan controles adicionales para evitar la reidentificación?	Sí
¿El resultado final se considera datos anonimizados?	Sí
***Pasos aplicables**	1, 2, 3, 4, 5

Pasos aplicables*



Caso de uso E	
Caso de uso	**Datos sintéticos**[8] para fines de desarrollo y prueba de aplicaciones, donde no se requiere la replicación de las características estadísticas de los datos originales (por ejemplo, utilizados para pruebas por proveedores subcontratados contratados para desarrollar y probar aplicaciones de nómina).
Descripción	Los datos sintéticos a nivel de registro se pueden crear a partir de los datos originales anonimizando en gran medida todos los atributos de datos utilizando las técnicas de anonimización de esta guía, de modo que todos los atributos de datos se modifican de manera muy significativa y todos los registros creados no coinciden con el registro de ningún individuo en los datos originales. En este caso, la aplicación de técnicas de anonimización no conservaría las características estadísticas de los datos originales y, por lo tanto, no es adecuada para fines sofisticados, como el entrenamiento de modelos de IA o el análisis de datos.
¿Se necesitan controles adicionales para evitar la reidentificación?	No[9]
¿El resultado final se considera datos anonimizados?	Sí
***Pasos aplicables**	1, 2, 3

Nota: En los datos sintéticos, los identificadores directos "falsos" utilizados no deben estar relacionados con una persona real, es decir, un documento nacional de identidad generado aleatoriamente con un nombre generado aleatoriamente no debe ser el mismo que el documento nacional de identidad y la combinación de nombres de una persona real.

8 Otro enfoque que no se aborda en esta guía es crear datos sintéticos desde cero. Esto se puede hacer generando aleatoriamente un conjunto de datos que simplemente cumpla con los requisitos de formato de datos, o generando un conjunto de datos que también conserve las características estadísticas del conjunto de datos original utilizando Inteligencia Artificial u otros métodos.

9 Consulte los supuestos en el Paso 5: Administrar sus riesgos de reidentificación y divulgación.

PASO 1: CONOZCA SUS DATOS

Aplicable a:

- Intercambio interno de datos (datos desidentificados)
- Retención de datos a largo plazo
- Intercambio interno de datos (datos anonimizados) o Intercambio externo de datos
- Datos sintéticos

Un registro de datos personales se compone de atributos de datos que tienen diversos grados de identificabilidad y sensibilidad a un individuo.

La anonimización generalmente implica la eliminación de identificadores directos y la modificación de identificadores indirectos. Los atributos objetivo generalmente se dejan sin cambios, excepto cuando el propósito es crear datos sintéticos. 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 objetivo
Clasificación de atributos de datos en un conjunto de datos	Estos son atributos de datos que son exclusivos de un individuo y se pueden usar como atributos de datos clave para volver a identificar a un individuo.	Estos son atributos de datos que no son exclusivos de un individuo, pero pueden volver a identificar a un individuo cuando se combinan con otra información (por ejemplo, una combinación de edad, sexo y código postal).	Estos son atributos de datos que contienen la utilidad principal del conjunto de datos. En el contexto de la evaluación de la adecuación de la anonimización, este atributo de datos puede ser de naturaleza sensible y puede dar lugar a un alto potencial de efecto adverso para un individuo cuando se divulga.

	Identificadores directos	Identificadores indirectos o seudoindentificadores	Atributos objetivo
Accesibilidad de los datos	Estos atributos de datos suelen ser públicos o de fácil acceso.	Estos atributos de datos pueden ser públicos o de fácil acceso.	Estos atributos de datos generalmente no son públicos ni de fácil acceso. No se pueden utilizar para la reidentificación, ya que suelen ser propietarios.
Ejemplos comunes en un conjunto de datos	• Nombre • Dirección de correo electrónico • Número de teléfono móvil • Número DNI • Número de pasaporte • Número de cuenta • Número de certificado de nacimiento • Número de permiso de trabajo • Nombre de usuario de redes sociales	• Edad • Género • Carrera • Fecha de nacimiento • Dirección • Código postal • Título del trabajo • Nombre de la empresa • Estado civil • 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)	• Transacciones (por ejemplo, compras) • Salario • Calificación crediticia • Póliza de seguro • Diagnóstico médico • Estado de vacunación

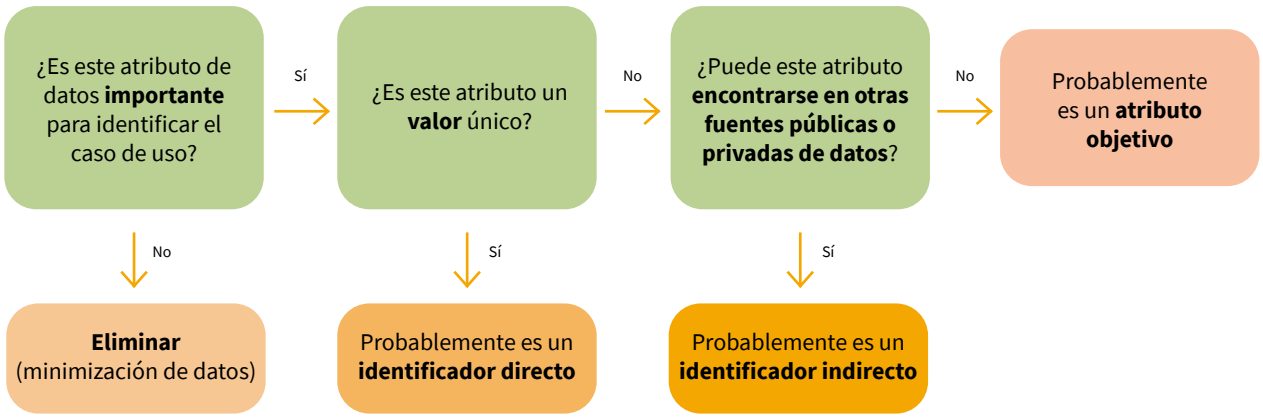
Ejemplo 1: Clasificación de atributos de datos en un registro de datos de empleados

Nº personal	Nombre	Departamento	Género	Fecha de nacimiento	Fecha de incorporación	Tipo de jornada
39192	Sandra García	Investigación & Desarrollo	M	08/01/1971	02/03/1997	Media jornada
37030	Paula Martínez	Ingeniería	M	15/05/1976	08/03/2015	Jornada completa
22722	Bernardo Sánchez	Ingeniería	H	31/12/1973	30/07/1991	Jornada completa
28760	Estefanía Gómez	Ingeniería	M	24/12/1970	18/03/2010	Media jornada
13902	Javier Muñoz	Recursos Humanos	H	15/07/1973	28/05/2012	Media jornada
Identificadores directos		Identificadores indirectos			Atributos objetivo	

Ejemplo 2: Clasificación de atributos de datos en un registro de datos de cliente

Nº personal	Nombre	Género	Fecha de nacimiento	Código postal	Ocupación	Ingresos	Educación	Estado civil
39192	Sandra García	M	05/08/1975	570150	Científico de datos	53.000€	Master	Viuda
37030	Paula Martínez	H	14/12/1973	787589	Profesor universidad	52.000€	Doctorado	Casado
22722	Bernardo Sánchez	H	02/03/1985	408600	Investigador	47.000€	Doctorado	Divorciado
28760	Estefanía Gómez	M	27/03/1968	570150	Administrador	48.000€	Licenciado	Casada
13902	Javier Muñoz	H	25/06/1967	199588	Arquitecto	50.000€	Master	Soltero
Identificadores directos		Identificadores indirectos			Atributos objetivo	Identificadores indirectos		

Cualquier atributo de datos que no sea necesario en el conjunto de datos resultante debe eliminarse como parte de la minimización de datos. A continuación, se proporciona un diagrama de flujo simple como ayuda para clasificar sus atributos de datos de manera adecuada.



PASO 2: DESIDENTIFICAR SUS DATOS

Aplicable a:

- Intercambio interno de datos (datos desidentificados)
- Retención de datos a largo plazo
- Intercambio interno de datos (datos anonimizados) o Intercambio externo de datos
- Datos sintéticos

Este paso siempre se realiza como parte del proceso de anonimización.

Primero, elimine todos los identificadores directos. En el ejemplo siguiente, se quitan todos los nombres. Cuando el conjunto de datos incluya otros identificadores directos, como el número documento nacional de identidad y la dirección de correo electrónico, también deben eliminarse.

~~Nombre~~	Edad	Serie favorita
~~Alex~~	25	The Big Bang Theory
~~Bosco~~	54	Friends
~~Charlene~~	42	Grey's Anatomy

Opcionalmente, asigne un seudónimo a cada registro si es necesario vincular el registro a un individuo único o al registro original para casos de uso como:

a) Fusión de datos

b) Análisis de múltiples registros relacionados con individuos únicos

c) Generación de conjuntos de datos sintéticos donde se requieren valores de identificador directo para el desarrollo y prueba de aplicaciones. Para este caso de uso, reemplace todos los identificadores directos necesarios con seudónimos

Los seudónimos deben ser únicos para cada identificador directo único (como se ilustra a continuación). La asignación de seudónimos también debe ser robusta (es decir, no ser reversible por partes no autorizadas a través de la deducción o el cálculo de los valores originales del identificador directo a partir de los seudónimos).

~~Nombre~~	Token	Edad	Serie favorita
~~Alex~~	1234	25	The Big Bang Theory
~~Bosco~~	5678	54	Friends
~~Charlene~~	5432	42	Grey's Anatomy

Si desea conservar la capacidad de vincular el registro de datos desidentificados al registro original en un momento posterior, deberá mantener la asignación entre los identificadores directos y los seudónimos. La tabla de asignación de identidad (que se ilustra a continuación) debe mantenerse de forma segura, ya que permite la reidentificación.

Nombre	Token
Alex	1234
Bosco	5678
Charlene	5432

PASO 3: APLICAR TÉCNICAS DE ANONIMIZACIÓN

Aplicable a:

- ✗ Intercambio interno de datos (datos desidentificados)
- ✓ Retención de datos a largo plazo
- ✓ Intercambio interno de datos (datos anonimizados) o Intercambio externo de datos
- ✓ Datos sintéticos

En este paso, aplicará técnicas de anonimización a los identificadores indirectos para que no se puedan combinar fácilmente con otros conjuntos de datos que puedan contener información adicional para volver a identificar a las personas. Para el caso de uso de datos sintéticos, también deben aplicarse técnicas de anonimización a los atributos objetivo.

Tenga en cuenta que la aplicación de estas técnicas modificará los valores de los datos y puede afectar a la utilidad de los datos anonimizados para algunos casos de uso (por ejemplo, análisis de datos). Las técnicas de anonimización recomendadas a continuación tienen en cuenta la utilidad potencial requerida para los datos a nivel de registro en cada caso de uso. Las organizaciones pueden utilizar otras técnicas de anonimización más allá de lo recomendado, si es relevante para su caso de uso.

Caso de uso	Técnicas de anonimización sugeridas para datos a nivel de registro
Intercambio interno de datos (datos anonimizados) o intercambio externo de datos	• **Supresión de registros:** la eliminación de un registro (es decir, una fila de datos, especialmente cuando dichos datos pueden contener valores de datos únicos que no se pueden anonimizar más). • **Supresión de atributos:** la eliminación de un atributo de datos (es decir, una columna de datos, especialmente cuando dichos datos no son necesarios en el conjunto de datos y pueden contener valores de datos únicos que no se pueden anonimizar más). • **Enmascaramiento de caracteres:** la sustitución de algunos caracteres del valor de datos por un símbolo coherente (por ejemplo, * o x). Por ejemplo, enmascarar un código postal implicaría cambiarlo de “28029” a “28xxx”. • **Generalización:** La reducción de la granularidad de los datos (por ejemplo, mediante la conversión de la edad de una persona en un rango de edad). Por ejemplo, generalizar la edad de una persona de “26 años” a “25-29 años”.

Caso de uso	Técnicas de anonimización sugeridas para datos a nivel de registro
Retención de datos a largo plazo	• **Supresión de registros o atributos** • **Enmascaramiento de caracteres** • **Generalización** • **Perturbación de datos:** Modificación de los valores en los datos agregando "ruido" a los datos originales (por ejemplo, valores aleatorios +/- a los datos). El grado de perturbación debe ser proporcional al rango de valores del atributo. Por ejemplo, la perturbación de los datos implicaría modificar los datos salariales de un individuo de "$ 256,654" a "$ 260,000" redondeando los datos a los $ 10,000 más cercanos. Alternativamente, el salario del individuo se puede modificar a "$ 250,554" restando un número aleatorio dentro de $ 10,000 de su valor original. *Nota: La agregación de datos también se puede realizar para este caso de uso cuando no se requieren datos a nivel de registro (consulte el Anexo A para obtener un ejemplo).*
Datos sintéticos	Aplicando una anonimización severa a los datos originales para crear datos sintéticos de modo que todos los atributos de datos (incluidos los atributos objetivo) se modifiquen significativamente. El conjunto de datos resultante y los registros individuales creados con esta metodología no tendrán ninguna semejanza con el registro de ningún individuo y no conservan las características del conjunto de datos original. Debido a la falta de semejanza del conjunto de datos resultante con el original, es adecuado para el desarrollo / prueba de aplicaciones, pero no para el entrenamiento de modelos de IA. • **Perturbación de datos** • **Intercambio:** reorganización de los datos en el conjunto de datos de forma aleatoria de modo que los valores de los atributos individuales todavía se representan en el conjunto de datos, pero generalmente no corresponden a los registros originales.

Consulte el anexo A para obtener más información sobre las diversas técnicas de anonimización y cómo aplicarlas. Consulte el anexo B para conocer las técnicas de anonimización sugeridas para aplicar en una lista de atributos de datos comunes.

Siguientes pasos: Después de aplicar las técnicas de anonimización apropiadas, continúe con el paso 4 para evaluar el nivel de riesgo. Repita los pasos 3 y 4 hasta que logre un valor de k-anonimidad de 3, 5 o más.

Nota: En los datos sintéticos, los identificadores directos "falsos" utilizados no deben estar relacionados con una persona real, es decir, un documento nacional de identidad generado aleatoriamente con un nombre generado aleatoriamente no debe ser el mismo que el documento nacional de identidad y la combinación de nombres de una persona real.

PASO 4: CALCULE SU RIESGO

Aplicable a:

- ✗ Intercambio interno de datos (datos desidentificados)
- ✓ Intercambio interno de datos (datos anonimizados) o Intercambio externo de datos
- ✓ Retención de datos a largo plazo
- ✗ Datos sintéticos

k-anonimidad[10] es un método fácil[11,12], para calcular el nivel de riesgo de reidentificación de un conjunto de datos. Básicamente se refiere al menor número de registros idénticos que se pueden agrupar en un conjunto de datos. Por lo general, se considera que el grupo más pequeño representa el peor de los casos al evaluar el riesgo general de reidentificación del conjunto de datos. Un valor de k-anonimidad de 1 significa que el registro es único. En general, solo se consideran identificadores indirectos para el cálculo de la k-anonimidad[13].

Un valor de k-anonimidad más alto significa que existe un menor riesgo de reidentificación, mientras que un valor de k-anonimidad más bajo implica un mayor riesgo. **En general, el umbral de la industria para el valor de k-anonimidad es de 3 o 5**[14] . Siempre que sea posible, debe establecerse un valor umbral de k-anonimidad más alto para minimizar cualquier riesgo de reidentificación.

En el Capítulo 3 (Anonimización) de "las Directrices de asesoramiento del PDPC sobre la Ley de Protección de Datos Personales para Temas Seleccionados" puede encontrar criterios para determinar si los datos pueden considerarse suficientemente anonimizados.

10 Más información sobre k-anonimidad y cómo usarla para evaluar el riesgo de reidentificación puede consultarse en los anexos C y D.

11 La k-anonimidad puede no ser adecuada para todos los tipos de conjuntos de datos u otros casos de uso complejos (por ejemplo, datos longitudinales o transaccionales en los que los mismos identificadores indirectos pueden aparecer en varios registros). Algoritmos Especiales Únicos de Detección (SUDA) y µ-Argus son otros métodos/herramientas para evaluar el riesgo de los conjuntos de datos compartidos.

12 Una limitación conocida del uso de la k-anonimidad es la revelación de atributos a través de ataques de homogeneidad que pueden afrontarse utilizando variantes de la k-anonimidad como la l-diversity y t-closeness. Estos temas están fuera del alcance de esta guía.

13 Los identificadores directos deberían haberse eliminado en el paso 2 y los seudónimos no deberían incluirse en el cálculo; de lo contrario, cada registro sería único.

14 Referencia de El marco de toma de decisiones de desidentificación por la Oficina del Comisionado de Información de Australia, CSIRO y Data 61.

K=2 en general			
	Código postal	**Edad**	**Serie Favorita**
k=2	22xxxx	21-25	La Casa de Papel
	22xxxx	21-25	La Casa de Papel
k=4	10xxxx	41-45	Peaky Blinders
	10xxxx	41-45	Peaky Blinders
	10xxxx	41-45	Peaky Blinders
	10xxxx	41-45	Peaky Blinders
k=2	58xxxx	56-60	Juego de Tronos
	58xxxx	56-60	Juego de Tronos
	58xxxx	56-60	Juego de Tronos

El diagrama anterior ilustra un conjunto de datos con tres grupos de registros idénticos. El valor k de cada grupo oscila entre 2 y 4. En general, el valor de k-anonimidad del conjunto de datos es 2, lo que refleja el valor más bajo (mayor riesgo) dentro de todo el conjunto de datos.[15]

Siguiente paso: Si se alcanza el umbral de valor de k-anonimidad, continúe con el paso 5. Si el valor de k-anonimidad es inferior al umbral establecido, vuelva al paso 3 y repita.

Nota: Siempre que sea posible, debe establecer un valor de k-anonimidad más alto (por ejemplo, 5 o más) para el intercambio de datos externos, mientras que se puede establecer un valor más bajo (por ejemplo, 3) para el intercambio de datos internos o la retención de datos a largo plazo. Sin embargo, si no puede anonimizar sus datos para lograrlo, debe implementar medidas de seguridad más estrictas para garantizar que los datos anonimizados no se divulguen a partes no autorizadas y se mitiguen los riesgos de reidentificación. Alternativamente, puede contratar a expertos para que proporcionen métodos de evaluación alternativos para lograr riesgos de reidentificación equivalentes.

15 La guía adopta el enfoque más conservador de observar el riesgo máximo. También hay otros enfoques (por ejemplo, riesgo medio y riesgo medio estricto).

PASO 5: GESTIONE SUS RIESGOS DE REIDENTIFICACIÓN Y DIVULGACIÓN

Aplicable a:

- ✅ Intercambio interno de datos (datos desidentificados)
- ✅ Retención de datos a largo plazo
- ✅ Intercambio interno de datos (datos anonimizados) o Intercambio externo de datos
- ❌ Datos sintéticos

En general, es prudente adoptar las medidas adecuadas para proteger sus datos contra los riesgos de reidentificación y revelación. Esto es teniendo en cuenta los futuros avances tecnológicos, así como de los conjuntos de datos desconocidos que podrían usarse para coincidir con su conjunto de datos anónimo y permitir que la reidentificación se realice más fácilmente de lo esperado en el momento de la anonimización.

Como buena práctica, los detalles del proceso de anonimización, los parámetros utilizados y los controles también deben registrarse claramente para futuras consultas. Dicha documentación facilita la revisión, el mantenimiento, el ajuste fino y las auditorías. Tenga en cuenta que dicha documentación debe conservarse de forma segura, ya que la revelación de los parámetros puede facilitar la reidentificación y la revelación de los datos anonimizados.

Existen varios tipos de riesgos de reidentificación y revelación. A continuación, se explican algunas fundamentales que debe evaluar al revisar la suficiencia de las medidas de protección que se han implementado.

1. **Reidentificación (revelación de identidad):** Determinar, con un alto nivel de confianza, la identidad de un individuo descrito por un registro específico. Esto podría surgir de escenarios como la anonimización insuficiente, la reidentificación mediante vinculación o la inversión del seudónimo. Por ejemplo, un proceso de anonimización que crea seudónimos basados en un algoritmo fácilmente adivinable y reversible, como reemplazar “1” por “a”, “2” por “b”, etc.

2. **Revelación de atributos:** Determinar, con un alto nivel de confianza, que un atributo descrito en el conjunto de datos pertenece a un individuo específico, incluso si el registro del individuo no se puede distinguir. Tomemos, por ejemplo, un conjunto de datos que contiene registros anónimos de clientes de un cirujano estético en particular que revela que todos sus clientes menores de 30 años se han sometido a un procedimiento en particular. Si se sabe que un individuo en particular tiene 28 años y es cliente de este cirujano, entonces sabemos que este individuo se ha sometido al procedimiento en particular, incluso si el registro del individuo no se puede distinguir de otros en el conjunto de datos anónimo.

3. Revelación de inferencias: Hacer una inferencia, con un alto nivel de confianza, sobre un individuo, incluso si él o ella no está en el conjunto de datos por propiedades estadísticas del conjunto de datos. Por ejemplo, si un conjunto de datos publicado por un investigador médico revela que el 70% de las personas mayores de 75 años tienen una determinada afección médica, esta información podría inferirse sobre una persona que no está en el conjunto de datos.

En general, la mayoría de las técnicas tradicionales de anonimización tienen como objetivo proteger contra la reidentificación y no necesariamente otros tipos de riesgos de revelación.

En la tabla siguiente se explica cuándo se recomiendan medidas contra los riesgos de reidentificación y revelación. En los párrafos siguientes se describe un conjunto de medidas básicas de protección (controles técnicos, de proceso y legales) para los casos de uso.

Caso de uso	¿Necesita gestionar los riesgos de reidentificación y revelación de conjuntos de datos desidentificados o anonimizados?
Intercambio interno de datos (datos desidentificados)	Como solo se ha aplicado la desidentificación para conservar una alta utilidad de los datos, el riesgo de reidentificación y revelación de los datos es mayor. Por lo tanto, se requiere protección para el conjunto de datos desidentificado. Las tablas de correspondencia de identidad, si las hay, deben estar protegidas. En el caso de una brecha de datos, la aplicación de técnicas de desidentificación, el cómo se protege el conjunto de datos desidentificado y el cómo se asegura la tabla de correspondencia se consideraría parte de los mecanismos de protección implementados.
Intercambio interno de datos (datos anonimizados)	Para reducir los riesgos de reidentificación y revelación, la anonimización debe aplicarse a los datos para su intercambio interno, cuando sea necesario, en los siguientes casos. Son (a) donde no se requieren datos personales detallados, (b) donde se pueden compartir datos confidenciales o (c) donde se comparte un gran conjunto de datos con más de un departamento. Se requiere protección básica para el conjunto de datos anonimizado. Las tablas de correspondencia de identidad, si las hay, deben protegerse y no compartirse con los otros departamentos internos.
Uso compartido de datos externos	Se requiere protección básica para el conjunto de datos anonimizado. Las tablas de correspondencia de identidad, si las hay, deben protegerse y no compartirse externamente.

Caso de uso	¿Necesita gestionar los riesgos de reidentificación y revelación de conjuntos de datos desidentificados o anonimizados?
Retención de datos a largo plazo	Se requiere protección básica para el conjunto de datos anonimizado. Todas las tablas de correspondencia de identidad deben destruirse de forma segura.

Para el caso de uso de datos sintéticos, se supone que los riesgos de reidentificación son mínimos cuando la anonimización se aplica en gran medida a todos los identificadores indirectos y atributos objetivo, de modo que los registros no se parezcan al conjunto de datos original. Como tal, no se requiere ninguna protección adicional de este conjunto de datos.

Controles técnicos y de procesos: Debe implementar medidas técnicas de protección para gestionar el riesgo de reidentificación y revelación de datos desidentificados y anonimizados. En la siguiente tabla se sugieren algunas buenas prácticas.

Debe revisar estas buenas prácticas para determinar si son suficientes para proteger sus datos desidentificados/anonimizados en función del grado de anonimización aplicado, la sensibilidad de los datos desidentificados/anonimizados y el caso de uso. Puede consultar la guía del **"PDPC: Guide to Data Protection Practices for ICT Systems"** para obtener medidas de protección adicionales cuando corresponda.

En la tabla, "Y" significa que se recomienda que adopte el control técnico correspondiente y "N/A" significa que el control técnico particular no es aplicable a ese caso de uso.

	Control técnico	Intercambio interno de datos (datos desidentificados)	Intercambio interno de datos (datos anonimizados)	Intercambio externo de datos	Retención de datos a largo plazo
Control de acceso y contraseñas	Implemente el control de acceso a nivel de aplicación para restringir el acceso a los datos a un nivel de usuario. Nivel mínimo de complejidad de la contraseña (es decir, mínimo 12 caracteres alfanuméricos con una mezcla de mayúsculas, minúsculas, números y caracteres especiales).	Y	Y	Y	Y

	Control técnico	Intercambio interno de datos (datos desidentificados)	Intercambio interno de datos (datos anonimizados)	Intercambio externo de datos	Retención de datos a largo plazo
Control de acceso y contraseñas	Revise regularmente las cuentas de usuario para asegurarse de que todas las cuentas estén activas y que los derechos asignados sean necesarios (por ejemplo, eliminar cuentas de usuario cuando un usuario haya abandonado la organización o actualizar los derechos del usuario cuando haya cambiado su rol dentro de la organización).	Y	Y	Y	Y
Seguridad para dispositivos de almacenamiento/bases de datos	Proteja los equipos mediante la aplicación de contraseñas. Ejemplos serian incluir ingresar la contraseña durante el arranque, requerir el inicio de sesión en el sistema operativo, bloquear la pantalla después de un período de inactividad, etc.	Y	Y	Y	Y
	Cifre el conjunto de datos. Revise el método de cifrado (por ejemplo, algoritmo y longitud de clave) periódicamente asegurarse de que el método de encriptación es reconocido por la industria como relevante y seguro.	Y *(cuando los datos involucrados son de naturaleza sensible o conjuntos de datos más grandes se comparten con más de un departamento, pero la anonimización no se aplica al conjunto de datos).*	N/A	N/A *(cuando se considere que el riesgo de reidentificación es bajo (por ejemplo, el k-anonimidad es de 5 o más), no es necesario aplicar cifrado al conjunto de datos anonimizado).*	N/A
	Cifre las tablas de correspondencia de identidades. Las tablas de correspondencia de identidades deben estar protegidas y no compartirse en todos los casos de uso.	Y	Y	Y	N/A *(Las tablas de correspondencia de identidades deben eliminarse)*
	Comunique la clave de descifrado del conjunto de datos por separado al destinatario de los datos compartidos/ exportados.	Y	N/A	N/A	N/A

	Control de procesos	Intercambio interno de datos (datos desidentificados)	Intercambio interno de datos (datos anonimizados)	Intercambio externo de datos	Retención de datos a largo plazo
Gestión de indicencias	Desarrolle un plan de gestión de brechas de datos para responder a las brechas de datos y administrar la pérdida de conjuntos de datos de manera más efectiva. El plan también debe incluir cómo gestionar la pérdida de tablas o información que podría permitir revertir los datos desidentificados /anonimizados a su estado original, lo que resultaría en que los datos perdidos se vuelvan a identificar. Consulte a continuación para obtener más información sobre la gestión de incidentes.	Y	Y	Y	Y
Controles de gobierno interno	Mantenga un registro central de todos los datos compartidos desidentificados/anonimizados para garantizar que los datos compartidos combinados no den lugar a una nueva identificación de los datos desidentificados/anonimizados.	Y	Y	Y	N/A
	Realizar periódicamente revisiones de reidentificación de los datos desidentificados / anonimizados.	Y	Y	Y	Y
	Asegúrese de que el destinatario (individuo o departamento) y el propósito del uso de los datos desidentificados / anonimizados hayan sido aprobados por las autoridades pertinentes dentro de la organización.	Y	Y	N/A	N/A
	Prohibir que el destinatario autorizado (individuo o departamento) comparta datos desidentificados / anonimizados con partes no autorizadas o intente volver a identificar los datos sin la aprobación de las autoridades pertinentes dentro de la organización.	Y	Y	N/A	N/A

	Control de procesos	Intercambio interno de datos (datos desidentificados)	Intercambio interno de datos (datos anonimizados)	Intercambio externo de datos	Retención de datos a largo plazo
Controles de gobierno interno	Elimine regularmente los datos desidentificados / anonimizados dentro de la organización cuando su propósito se ha cumplido y ya no hay necesidad de los datos.	Y	Y	N/A	N/A
	Realizar periódicamente comprobaciones/auditorías internas para garantizar el cumplimiento de los procesos.	Y	Y	Y	Y

Gestión de incidentes: Las organizaciones deben identificar los riesgos de brechas de datos que impliquen una tabla de correspondencia de identidad, datos desidentificados y datos anónimos, e incorporar escenarios relevantes en sus planes de gestión de incidentes. Las siguientes consideraciones pueden ser relevantes para la notificación de brechas de datos y las investigaciones internas:

- **Pérdida de datos desidentificados y tabla de correspondencia de identidad**

Las brechas tanto de los datos desidentificados como de la tabla de gestión de identidad será similar a las brechas de los datos personales. En tal caso, la organización debe evaluar si una brecha de datos es notificable y notificarlo a las personas afectadas y/o a la Comisión, cuando se considere que es notificable en virtud de la obligación de notificación de brechas de datos.

- **Pérdida solamente de datos desidentificados**

Si los datos desidentificados se han filtrado externamente, es necesaria una evaluación. La organización debe evaluar si la brecha de datos es notificable, ya que los datos desidentificados tienen un mayor riesgo de reidentificación. Sin embargo, el uso de la desidentificación y otras salvaguardas para proteger los datos y la tabla de correspondencia de identidad podría considerarse parte de los mecanismos de protección implementados por la organización.

- **Pérdida de datos anonimizados y tabla de correspondencia de identidad**

Las organizaciones tienen que evaluar el riesgo de reidentificación. Cuando se determine que es elevada, las organizaciones deben determinar si la brecha de datos es notificable y notificarlo a las personas afectadas y/o a la Comisión, cuando se considere que es notificable en virtud de la obligación de notificación de brechas de datos.

- **Pérdida solamente de datos anonimizados**

Cuando la organización ha aplicado las técnicas de anonimización correctamente, no necesita informar de la filtración como una brecha de datos de notificación obligatoria. Sin embargo, aún debe investigar el incidente para comprender la causa para mejorar sus salvaguardas internas contra futuros incidentes de brechas de datos.

- **Pérdida solamente de la tabla de asignación de identidad**

Si los conjuntos de datos para los que se utilizó la tabla de correspondencia de identidad aún están protegidos, las organizaciones no necesitan informar de la infracción, ya que una tabla de correspondencia de identidad por sí sola no son datos personales. Sin embargo, la organización debe generar inmediatamente nuevos seudónimos para sus conjuntos de datos y una nueva tabla de correspondencia de identidad. También debe investigar el incidente para comprender la causa para mejorar sus salvaguardas internas contra futuros incidentes de brechas de datos.

Controles legales: Las organizaciones deben protegerse garantizando que los terceros destinatarios de sus datos anonimizados incorporen la protección pertinente a los datos anonimizados compartidos para minimizar los riesgos de reidentificación. Las buenas prácticas de la siguiente tabla se han tomado del documento del PDPC *"Trusted Data Sharing Framework"*.

	Control legal	Intercambio interno de datos (datos desidentificados)	Intercambio interno de datos (datos anonimizados)	Intercambio externo de datos	Retención de datos a largo plazo
Acuerdo de intercambio de datos	Asegúrese de que los datos solo se utilicen para fines permitidos (por ejemplo, no revelación a partes no autorizadas) y que se asigne responsabilidad por incumplimientos contractuales.	N/A	N/A	Y	N/A
	Prohibir que terceros destinatarios intenten volver a identificar los conjuntos de datos anónimos que se han compartido.	N/A	N/A	Y	N/A
	Asegúrese de que terceros destinatarios cumplan con la protección pertinente sobre los datos anónimos compartidos según los controles internos de la organización.	N/A	N/A	Y	N/A

ANEXO A: TÉCNICAS BÁSICAS DE ANONIMIZACIÓN DE DATOS

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.
Cuando usarlo	La supresión de registros se utiliza para eliminar registros atípicos que son únicos o que no cumplen otros criterios, como la k-anonimidad, del conjunto de datos anonimizado. Los valores atípicos pueden conducir a una fácil reidentificación. Se puede aplicar antes o después de que se hayan aplicado otras técnicas (por ejemplo, la generalización).
Cómo usarlo	Elimine todo el registro. Tenga en cuenta que la supresión debe ser permanente y no solo una función de "ocultar fila"[16] ; del mismo modo, la "redacción" puede no ser suficiente si los datos subyacentes siguen siendo accesibles.
Otros consejos	• Consulte el ejemplo de la sección sobre generalización para ilustrar cómo se utiliza la supresión de registros. • Tenga en cuenta que la eliminación de registros puede afectar al conjunto de datos (por ejemplo, en términos estadísticos como el promedio y la mediana).

16 Esto se refiere al uso de la función "ocultar fila" en su software de hoja de cálculo.

Enmascaramiento de caracteres	
Descripción	El enmascaramiento de caracteres se refiere al cambio de los caracteres de un valor de datos. Esto se puede hacer mediante el uso de un símbolo consistente (por ejemplo, "*" o "x"). Normalmente, el enmascaramiento se aplica solo a algunos caracteres del atributo.
Cuando usarlo	El enmascaramiento de caracteres se utiliza cuando el valor de los datos es una cadena de caracteres y ocultar parte de ella es suficiente para proporcionar el grado de anonimato requerido.
Cómo usarlo	En función de la naturaleza del atributo, reemplace los caracteres apropiados por un símbolo elegido. Dependiendo del tipo de atributo, puede decidir reemplazar un número fijo de caracteres (por ejemplo, para números de tarjetas de crédito) o un número variable de caracteres (por ejemplo, para la dirección de correo electrónico).
Otros consejos	• Tenga en cuenta que el enmascaramiento puede necesitar tener en cuenta si la longitud de los datos originales proporciona información sobre los datos originales. El conocimiento de la materia es fundamental, especialmente para el enmascaramiento parcial para garantizar que los caracteres correctos estén enmascarados. También se puede aplicar una consideración especial a las sumas de comprobación dentro de los datos; a veces, se puede usar una suma de comprobación para recuperar (otras partes de) los datos enmascarados. En cuanto al enmascaramiento completo, el atributo podría suprimirse alternativamente a menos que la longitud de los datos sea de cierta relevancia. • El escenario de enmascarar los datos de tal manera que los interesados estén destinados a reconocer sus propios datos es especial y no pertenece a los objetivos habituales de la anonimización de los datos. Un ejemplo de esto es la publicación de los resultados de los sorteos, donde los nombres y los números documento nacional de identidad parcialmente enmascarados de los ganadores del sorteo suelen publicarse para que las personas se reconozcan como ganadores. Otro ejemplo es información como el número de tarjeta de crédito de un individuo que se enmascara en una aplicación o una declaración dirigida al individuo. Tenga en cuenta que, en general, los datos anonimizados no deben ser reconocibles ni siquiera para el propio interesado.

Ejemplo.
Este ejemplo muestra una tienda de comida on-line que realiza un estudio de su demanda de entrega a partir de datos históricos para mejorar la eficiencia operativa. La compañía enmascaró los últimos 4 dígitos de los códigos postales, dejando los primeros 2 dígitos, que corresponden al "código de sector" dentro de Singapur.

Antes de la anonimización		
Código postal	**Franja horaria de entrega favorita**	**Número medio de pedidos al mes**
100111	8 pm a 9 pm	2
200222	11 am a 1 pm	8
300333	2 pm a 3 pm	1

Después del enmascaramiento parcial del código postal		
Código postal	**Franja horaria de entrega favorita**	**Número medio de pedidos al mes**
10xxxx	8 pm a 9 pm	2
20xxxx	11 am a 1 pm	8
30xxxx	2 pm a 3 pm	1

Seudonimización

Descripción	La seudonimización se refiere a la sustitución de datos de identificación por valores inventados. También se conoce como codificación. Los seudónimos pueden ser irreversibles cuando los valores originales se eliminan correctamente y la seudonimización se realiza de una manera no repetible. También pueden ser reversibles (por el propietario de los datos originales) cuando los valores originales se guardan de forma segura, pero se pueden recuperar y vincular al seudónimo en caso de que surja la necesidad.[17] Los seudónimos persistentes permiten la vinculación mediante el uso de los mismos valores de seudónimo para representar al mismo individuo en diferentes conjuntos de datos. Sin embargo, se pueden usar diferentes seudónimos para representar al mismo individuo en diferentes conjuntos de datos para evitar la vinculación de los diferentes conjuntos de datos. Los seudónimos también se pueden generar de forma aleatoria o determinista.
Cuando usarlo	La seudonimización se utiliza cuando los valores de los datos deben distinguirse de forma única y no se conserva ningún carácter o cualquier otra información implícita sobre los identificadores directos del atributo original.
Cómo usarlo	Reemplace los valores de atributo respectivos por valores inventados. Una forma de hacerlo es generar previamente una lista de valores inventados y seleccionar aleatoriamente de esta lista para reemplazar cada uno de los valores originales. Los valores inventados deben ser únicos y no deben tener relación con los valores originales (de modo que se puedan derivar los valores originales de los seudónimos).

17 Por ejemplo, en el caso de que un estudio de investigación arroje resultados que puedan proporcionar una advertencia útil a un sujeto de datos.

Seudonimización

Otros consejos

- Al asignar seudónimos, asegúrese de no reutilizar seudónimos que ya se hayan utilizado en el mismo conjunto de datos, especialmente cuando se generan aleatoriamente. Además, evite usar exactamente el mismo generador de seudónimos sobre varios atributos sin un cambio (por ejemplo, al menos use una secuencia aleatoria diferente).

  Los seudónimos persistentes generalmente proporcionan una mejor utilidad al mantener la integridad referencial en todos los conjuntos de datos.

  Para los seudónimos reversibles, la tabla de correspondencia de identidades no se puede compartir con el destinatario; debe conservarse de forma segura y solo puede ser utilizado por la organización cuando sea necesario volver a identificar a la(s) persona(s).
- Del mismo modo, si se utiliza el cifrado o una función hash para seudonimizar los datos, la clave de cifrado o el algoritmo hash y el valor de sal para el hash deben protegerse de forma segura contra el acceso no autorizado. Esto se debe a que una fuga de dicha información podría resultar en una brecha de datos al permitir la reversión del cifrado o usar tablas precalculadas para inferir los datos que se hashean (especialmente para datos que siguen formatos predeterminados, como en los documento nacional de identidad).

  Lo mismo se aplica a los generadores de números pseudoaleatorios, que requieren una semilla. La seguridad de cualquier clave utilizada debe garantizarse como con cualquier otro tipo de cifrado o proceso reversible. Las organizaciones también deben revisar periódicamente el método de cifrado (por ejemplo, algoritmo y longitud de clave) y la función hash para garantizar que la industria lo reconozca como método relevante y seguro.[18]

- En algunos casos, los seudónimos pueden necesitar seguir la estructura o el tipo de datos del valor original (por ejemplo, para que los seudónimos sean utilizables en aplicaciones de software); en tales casos, pueden ser necesarios generadores de seudónimos especiales para crear conjuntos de datos sintéticos o, en algunos casos, se puede considerar el llamado "cifrado de preservación de formato", que crea seudónimos que tienen el mismo formato que los datos originales.

18 Tenga en cuenta que confiar en un proceso de reversión propietario o "secreto" (con o sin una clave) tiene un mayor riesgo de ser decodificado y roto en comparación con el uso de un cifrado estándar basado en claves o hashing.

Ejemplo.
Este ejemplo muestra que la seudonimización se aplica a los nombres de las personas que obtuvieron sus permisos de conducir y a cierta información sobre ellos. En este ejemplo, los nombres se reemplazaron con seudónimos en lugar de suprimir el atributo porque la organización quería poder revertir la seudonimización si era necesario.

Antes de la anonimización		
Persona	**Resultado de la evaluación previa**	**Horas de clases tomadas antes de aprobar**
Jose Muñoz	A	20
Pedro Duran	B	26
Patricia Martínez	C	30
Marta Peña	D	29
Soraya García	B	32
Nicolás Pérez	A	25

Después de seudonimizar el atributo “Persona”		
Persona	**Resultado de la evaluación previa**	**Horas de clases tomadas antes de aprobar**
416765	A	20
562396	B	26
964825	C	30
873892	D	29
239976	B	32
943145	A	25

Para la seudonimización reversible, la tabla de correspondencia de identidad se mantiene de forma segura en caso de que exista una necesidad futura legítima de volver a identificar a las personas. Los controles de seguridad (incluidos los administrativos y técnicos) también deben utilizarse para proteger la tabla de correspondencia de identidades.

Tabla de asignación de identidades (codificación única):	
Seudónimo	**Persona**
416765	Jose Muñoz
562396	Pedro Duran
964825	Patricia Martínez
873892	Marta Peña
239976	Soraya García
943145	Nicolás Pérez

Para mayor seguridad con respecto a la tabla de correspondencia de identidades, se puede utilizar la codificación doble. A continuación del ejemplo anterior, este ejemplo muestra la tabla de correspondencia adicional, que se coloca con un tercero de confianza.

Con la doble codificación, la identidad de los individuos solo se puede conocer cuando tanto el tercero de confianza (que tiene la tabla de enlace) como la organización (que tiene la tabla de correspondencia de identidad) juntan sus datos.

Después de la anonimización		
Persona	**Resultado de la evaluación previa**	**Horas de clases tomadas antes de aprobar**
373666	A	20
594824	B	26
839933	C	30
280074	D	29
746791	B	32
785282	A	25

Tabla de enlace (mantenida de forma segura solo por un tercero de confianza e incluso la organización la eliminará eventualmente. El tercero no recibe ninguna otra información):	
Seudónimo	**Seudónimo provisional**
373666	OQCPBL
594824	ALGKTY
839933	CGFFNF
280074	BZMHCP
746791	RTJYGR
785282	RCNVJD

Tabla de asignación de identidad (mantenida de forma segura por la organización)	
Seudónimo provisional	**Persona**
OQCPBL	Jose Muñoz
ALGKTY	Pedro Duran
CGFFNF	Patricia Martínez
BZMHCP	Marta Peña
RTJYGR	Soraya García
RCNVJD	Nicolás Pérez

Nota: Tanto en la tabla de vinculación como en la tabla de correspondencia de identidades, es una buena práctica codificar el orden de los registros en lugar de dejarlo en el mismo orden que el conjunto de datos. En este ejemplo, los registros de ambas tablas se dejan en el orden original para facilitar la visualización.

Generalización	
Descripción	La generalización es una reducción deliberada de la precisión de los datos. Los ejemplos incluyen convertir la edad de una persona en un rango de edad o una ubicación precisa en una ubicación menos precisa. Esta técnica también se conoce como recodificación.
Cuando usarlo	La generalización se utiliza para valores que pueden generalizarse y seguir siendo útiles para el propósito previsto.
Cómo usarlo	Diseñe categorías de datos y reglas apropiadas para traducir datos. Considere la posibilidad de suprimir cualquier registro que aún se destaque después de la traducción (es decir, la generalización).
Otros consejos	• Elija un intervalo de datos adecuado. Un rango de datos que es demasiado grande puede significar una pérdida significativa en la utilidad de datos, mientras que un rango de datos que es demasiado pequeño puede significar que los datos apenas se modifican y, por lo tanto, aún son fáciles de volver a identificar. Si se utiliza la k-anonimidad, el valor k elegido también afectará al rango de datos. Tenga en cuenta que el primer y el último rango pueden ser un rango más grande para acomodar el número típicamente menor de registros en estos extremos; esto a menudo se conoce como codificación superior / inferior.

Ejemplo.
En este ejemplo, el conjunto de datos contiene el nombre de la persona (que ya ha sido seudonimizado), su edad en años y la dirección residencial.

Antes de la anonimización			
Número de serie	**Persona**	**Edad**	**Dirección**
1	357703	24	Avenida de Madrid, 22 1ºB
2	233121	31	Calle Alcalá, 18 1ºB
3	938637	44	Calle Mayor, 27 3ºB
4	591493	29	Avenida de Madrid, 22 6ºC
5	202626	23	Calle Serrano, 40 3ºD
6	888948	75	Carretera de Stonehenge, 5
7	175878	28	Calle Serrano, 40 5ºA
8	312304	50	Calle Mayor, 27 1ºA
9	214025	30	Avenida de Madrid, 22 2ºA
10	271714	37	Alcalá, 18 1ºC
11	341338	22	Calle Serrano, 40 7ºA
12	529057	25	Calle Serrano, 40 2ºB
13	390438	39	Calle Alcalá, 18 4ºB

Para el atributo “Edad”, el enfoque adoptado es generalizar en los siguientes rangos de edad.

< 20	21 -30	31 - 40	41 - 50	51 - 60	> 60

Para la “Dirección”, un enfoque posible es eliminar el número de bloque / casa y conservar solo el nombre de la carretera.

Después de la generalización de los atributos “Edad” y “Dirección”:			
Número de serie	**Persona**	**Edad**	**Dirección**
1	357703	21-30	Avenida de Madrid, 22 1ºB
2	233121	31-40	Calle Alcalá, 18 1ºB
3	938637	41-50	Calle Mayor, 27 3ºB
4	591493	21-30	Avenida de Madrid, 22 6ºC
5	202626	21-30	Calle Serrano, 40 3ºD
6	888948	>60	Carretera de Stonehenge, 5
7	175878	21-30	Calle Serrano, 40 5ºA
8	312304	41-50	Calle Mayor, 27 1ºA
9	214025	21-30	Avenida de Madrid, 22 2ºA
10	271714	31-40	Alcalá, 18 1ºC
11	341338	21-30	Calle Serrano, 40 7ºA
12	529057	21-30	Calle Serrano, 40 2ºB
13	390438	31-40	Calle Alcalá, 18 4ºB

Como ejemplo, supongamos que hay, de hecho, solo una unidad residencial en Carretera de Stonehenge, 5. La dirección exacta se puede derivar a pesar de que los datos han pasado por la generalización. Esto podría considerarse “demasiado único”.

Por lo tanto, como siguiente paso de la generalización, el registro 6 podría eliminarse (es decir, utilizando la técnica de supresión de registros) ya que la dirección sigue siendo “demasiado única” después de eliminar el número de unidad. Alternativamente, todas las direcciones podrían generalizarse en mayor medida (por ejemplo, ciudad o distrito) de modo que no sea necesaria la supresión. Sin embargo, esto puede afectar a la utilidad de los datos mucho más que suprimir algunos registros del conjunto de datos.

Intercambio	
Descripción	El propósito del intercambio es reorganizar los datos en el conjunto de datos de modo que los valores de los atributos individuales sigan representados en el conjunto de datos, pero generalmente no correspondan a los registros originales. Esta técnica también se conoce como barajado y permutación.
Cuando usarlo	El intercambio se utiliza cuando el análisis posterior solo necesita mirar datos agregados o el análisis es a nivel intra-atributo; en otras palabras, no hay necesidad de analizar las relaciones entre los atributos a nivel de registro.
Cómo usarlo	Primero, identifique qué atributos intercambiar. A continuación, para cada valor del atributo, intercambie o reasigne el valor a otros registros del conjunto de datos.
Otros consejos	• Evalúe y decida qué atributos (columnas) deben intercambiarse. Dependiendo de la situación, las organizaciones pueden decidir que, por ejemplo, solo los atributos (columnas) que contienen valores que son relativamente identificables deben intercambiarse.

Ejemplo.
En este ejemplo, el conjunto de datos contiene información sobre los registros de clientes de una organización empresarial.

Antes de la anonimización				
Persona	**Título del trabajo**	**Fecha de nacimiento**	**Tipo de membresía**	**Promedio de visitas por mes**
A	Profesor universitario	03/01/1970	Plata	0
B	Vendedor	05/02/1972	Platino	5
C	Abogado	07/03/1985	Oro	2
D	Profesional de TI	10/04/1990	Plata	1
E	Enfermera	13/05/1995	Plata	2

Después de la anonimización*				
Persona	**Título del trabajo**	**Fecha de nacimiento**	**Tipo de membresía**	**Promedio de visitas por mes**
A	Abogado	10/04/1990	Plata	1
B	Enfermera	07/03/1985	Plata	2
C	Vendedor	13/05/1995	Platino	5
D	Profesional de TI	03/01/1970	Plata	2
E	Profesor universitario	05/02/1972	Oro	0

**En este ejemplo, se han intercambiado todos los valores de todos los atributos.*

Nota: Por otro lado, si el propósito del conjunto de datos anonimizado es estudiar las relaciones entre el perfil del trabajo y los patrones de consumo, otros métodos de anonimización pueden ser más adecuados (por ejemplo, la generalización de los títulos de trabajo, lo que podría dar lugar a que el "profesor universitario" se modifique para convertirse en "docencia").

Perturbación de datos	
Descripción	Los valores del conjunto de datos original se modifican para que sean ligeramente diferentes.
Cuando usarlo	La perturbación de datos se utiliza para identificadores indirectos (normalmente números y fechas), que pueden ser potencialmente identificables cuando se combinan con otras fuentes de datos, pero los cambios leves en el valor son aceptables para el atributo. Esta técnica no debe utilizarse cuando la precisión de los datos sea crucial.
Cómo usarlo	Depende de la técnica exacta de perturbación de datos utilizada. Estos incluyen redondear y agregar ruido aleatorio. El ejemplo de esta sección muestra el redondeo de base x.
Otros consejos	• El grado de perturbación debe ser proporcional al rango de valores del atributo. Si la base es demasiado pequeña, el efecto de anonimización será más débil; por otro lado, si la base es demasiado grande, los valores finales serán demasiado diferentes del original y la utilidad del conjunto de datos probablemente se reducirá. • Tenga en cuenta que cuando el cálculo se realiza sobre valores de atributos que se han perturbado antes, el valor resultante puede experimentar perturbaciones en una medida aún mayor.

Ejemplo.
En este ejemplo, el conjunto de datos contiene información que se utilizará para la investigación sobre el posible vínculo entre la altura, el peso y la edad de una persona, si la persona fuma y si la persona tiene "enfermedad A" y / o "enfermedad B". El nombre de la persona ya ha sido seudonimizado.

A continuación, se aplica el siguiente redondeo:

Atributo	Anonimación técnica
Altura (en cm)	Redondeo base-5 (se elige 5, siendo algo proporcional al valor de altura típico de 120 a 190 cm).
Peso (en kg)	Redondeo base-3 (se elige 3, siendo algo proporcional al valor de peso típico de 40 a 100 kg).
Edad (en años)	Redondeo de base 3 (se elige 3, siendo algo proporcional al valor de edad típico de 10 a 100 años).
Los atributos restantes	Nil, porque no son numéricos y difíciles de modificar sin un cambio sustancial en el valor.

Conjunto de datos antes de la anonimización						
Persona	**Altura**	**Peso (kg)**	**Edad (años)**	**¿Fuma?**	**¿Enfermedad A?**	**¿Enfermedad B?**
198740	160	50	30	No	No	No
287402	177	70	36	No	No	Sí
398747	158	46	20	Sí	Sí	No
498732	173	75	22	No	No	No
598772	169	82	44	Sí	Sí	Sí

Conjunto de datos después de la anonimización						
Persona	**Altura**	**Peso (kg)**	**Edad (años)**	**¿Fuma?**	**¿Enfermedad A?**	**¿Enfermedad B?**
198740	160	51	30	No	No	No
287402	175	69	36	No	No	Sí
398747	160	45	18	Sí	Sí	No
498732	175	75	21	No	No	No
598772	170	81	42	Sí	Sí	Sí

Nota: Para el redondeo base-x, los valores de atributo que se van a redondear se redondean al múltiplo más cercano de x.

Agregación de datos	
Descripción	La agregación de datos se refiere a la conversión de un conjunto de datos de una lista de registros a valores resumidos.
Cuando usarlo	Se utiliza cuando no se requieren registros individuales y los datos agregados son suficientes para el propósito.
Cómo usarlo	Una discusión detallada de las medidas estadísticas está más allá del alcance de esta guía, sin embargo, las formas típicas incluyen el uso de totales o promedios, etc. También puede ser útil discutir con el destinatario de los datos sobre la utilidad esperada y encontrar un compromiso adecuado.
Otros consejos	• Cuando corresponda, tenga cuidado con los grupos que tienen muy pocos registros después de realizar la agregación. En el siguiente ejemplo, si los datos agregados incluyen un solo registro en cualquiera de las categorías, podría ser fácil para alguien con algún conocimiento adicional identificar a un donante. • Por lo tanto, la agregación puede necesitar ser aplicada en combinación con la supresión. Es posible que sea necesario eliminar algunos atributos, ya que contienen detalles que no se pueden agregar y es posible que sea necesario agregar nuevos atributos (por ejemplo, para contener los valores agregados recién calculados).

Ejemplo.
En este ejemplo, una organización benéfica tiene registros de las donaciones realizadas, así como cierta información sobre los donantes. La organización benéfica evaluó que los datos agregados son suficientes para que un consultor externo realice análisis de datos, por lo tanto, realizó la agregación de datos en el conjunto de datos original.

Conjunto de datos original		
Donante	**Ingesos mensuales ($)**	**Cantidad donada en 2016 ($)**
Donante A	4000	210
Donante B	4900	420
Donante C	2200	150
Donante D	4200	110
Donante E	5500	260
Donante F	2600	40
Donante G	3300	130
Donante H	5500	210
Donante I	1600	380
Donante J	3200	80
Donante K	2000	440
Donante L	5800	400
Donante M	4600	390
Donante N	1900	480
Donante O	1700	320
Donante P	2400	330
Donante Q	4300	390

Conjunto de datos original		
Donante	**Ingesos mensuales ($)**	**Cantidad donada en 2016 ($)**
Donante R	2300	260
Donante S	3500	80
Donante T	1700	290

Conjunto de datos anonimizado		
Ingesos mensuales ($)	**Número de donaciones recibidas (2016)**	**Suma de la cantidad donada en 2016 ($)**
1000-1999	4	1470
2000-2999	5	1220
3000-3999	3	290
4000-4999	5	1520
5000-6000	3	870
TOTAL	**20**	**5370**

ANEXO B: ATRIBUTOS COMUNES DE LOS DATOS Y TÉCNICAS DE ANONIMIZACIÓN SUGERIDAS

Identificadores directos

La siguiente tabla proporciona sugerencias sobre técnicas de anonimización que se pueden aplicar a algunos tipos comunes de identificadores directos. En general, los identificadores directos deben suprimirse (eliminarse) o seudonimizarse. Si se requiere la asignación de seudónimos, generalmente es suficiente un conjunto (es decir, una columna) de seudónimos por conjunto de datos.

Para el caso de uso de datos sintéticos, todas las columnas de identificador directo se pueden conservar, pero deben reemplazarse con valores seudonimizados.

Supresión de registros	Técnica de uso común	Ejemplo	
		Antes	Después
• Nombre • Dirección de correo electrónico • Número de teléfono móvil • Número documento nacional de identidad • Pasaporte número • Número de cuenta • Número de certificado de nacimiento • Número de identificación extranjero (FIN) • Número de permiso de trabajo	**Supresión de atributos**	John Tan	(Suprimido)
	Asignación de seudónimos, por ejemplo:		
	• Reemplace los valores de identificador directo con valores aleatorios únicos; o	John Tan	123456
	• Reemplace los valores de identificador directo por valores generados aleatoriamente que sigan el formato de los datos.	John.tan@gmail.com S8822311H	123456@abc.com S8512345A

Identificadores indirectos

La siguiente tabla proporciona sugerencias sobre técnicas de anonimización que se pueden aplicar a algunos tipos comunes de identificadores indirectos. Debe optar por aplicar una o más de las técnicas a cada identificador indirecto (por ejemplo, aplicar la generalización y el intercambio a la edad, según su caso de uso).

Para el caso de uso de datos sintéticos, dos técnicas útiles son el intercambio de datos y la perturbación de datos. Estos se aplican a todos los identificadores indirectos.

Identificadores indirectos	Técnica(s) de uso común	Ejemplo(s)	
		Antes	Después
• Edad • Altura • Peso	**Generalización:** Generalizar la edad/altura/peso a rangos de 5 o 10 años/cm/kg.	Registro #1: 24 Registro #2: 39 Registro #3: 18	**Generalización:** (rango de edad de 5 años): Registro #1: 21 a 25 Registro #2: 36 a 40 Registro #3: 16 a 20
	Perturbación de datos: Agregue valores aleatorios (+/- 5) al valor original.		**Perturbación de datos:** Registro #1: 25 Registro #2: 36 Registro #3: 17
	Intercambio: Cambie aleatoriamente la edad/altura / peso asociado con cada registro.		**Intercambio:** Registro #1: 39 Registro #2: 18 Registro #3: 24
• Género	Este atributo de datos indirectos generalmente solo tiene dos valores genéricos no identificativos: M o F y, por lo tanto, generalmente es seguro retenerlo tal como está. Para el caso de uso de datos sintéticos, la siguiente técnica se puede aplicar a este atributo. **Intercambio:** Cambie aleatoriamente el género dentro del conjunto de datos.	Registro #1: M Registro #2: M Registro #3: F Registro #4: M	**Intercambio:** Registro #1: M Registro #2: F Registro #3: M Registro #4: M

Identificadores indirectos	Técnica(s) de uso común	Ejemplo(s)	
		Antes	Después
• Raza • Estado civil	**Generalización:** Dependiendo de su conjunto de datos, puede combinar y generalizar grupos étnicos o estados civiles seleccionados en una categoría etiquetada como "Otros". Esto debe hacerse si hay grupos étnicos / estados civiles únicos o muy pocos de los mismos grupos étnicos / estados civiles dentro de su conjunto de datos. **Intercambio:** Cambie aleatoriamente la raza o el estado civil dentro del conjunto de datos.	Registro #1: Indio Registro #2: Chino Registro #3: Chino Registro #4: Malayo Registro #5: Euroasiático	**Generalización:** Registro #1: Otros Registro #2: Chino Registro #3: Chino Registro #4: Otros Registro #5: Otros **Intercambio:** Registro #1: Malayo Registro #2: Chino Registro #3: Indio Registro #4: Euroasiático Registro #5: Chino
• Fecha de nacimiento	**Generalización:** Generalice la fecha de nacimiento al año, o mes y año.	Registro #1: 1 Feb 2003 Registro #2: 15 Aug 1990 Registro #3: 30 Dec 1998	**Generalización** (mes y año): Registro #1: Feb 2003 Registro #2: Agosto 1990 Registro #3: Dic 1998
	Perturbación de datos: Modifique aleatoriamente la fecha (por ejemplo, +/- 30 días a partir de la fecha original).		**Perturbación de datos:** Registro #1: 20 Jan 2003 Registro #2: 18 Aug 1990 Registro #3: 6 Jan 1999
	Intercambio: Cambie aleatoriamente las fechas dentro del conjunto de datos.		**Intercambio:** Registro #1: 30 Dic 1998 Registro #2: 1 Feb 2003 Registro #3: 15 Aug 1990

Identificadores indirectos	Técnica(s) de uso común	Ejemplo(s)	
		Antes	Después
• Dirección	**Generalización:** Generalizar la dirección a las zonas predefinidas (por ejemplo, con referencia a un plan de ordenación urbana en distritos	Distrito Salamanca, 28006 Madrid	**Intercambio:** Registro #1: Calle Mayor, 27 3ºB Registro #2: Calle Serrano, 40 3ºD
	Intercambio: Cambie aleatoriamente las direcciones dentro del conjunto de datos. Nota: Para las direcciones, los números de unidad pueden ser identificativos. Cuando no sea necesario, los números de unidad deben eliminarse del conjunto de datos.	Registro #1: Calle Serrano, 40 3ºD Registro #2: Calle Mayor, 27 3ºB	
• Código postal	**Enmascaramiento de caracteres:** Enmascara los últimos cuatro dígitos del código postal.	117438	**Enmascaramiento de caracteres:** 11xxxx
	Intercambio: Cambie aleatoriamente los códigos postales dentro del conjunto de datos.	Registro #1: 117438 Registro #2: 828755	**Intercambio:** Registro#1: 828755 Registro#2: 117438

Identificadores indirectos	Técnica(s) de uso común	Ejemplo(s)	
		Antes	Después
• Título del trabajo	**Generalización:** No hay una manera fácil de anonimizar los títulos de trabajo de una manera automatizada porque los títulos de trabajo no son estándar, y las organizaciones pueden inventar los suyos propios. Una forma es generalizar los títulos de trabajo a una taxonomía predefinida de la naturaleza del trabajo y / o los niveles de trabajo. Sin embargo, es probable que la correspondencia tenga que hacerse manualmente. **Intercambio:** Cambie aleatoriamente los títulos de trabajo dentro del conjunto de datos.	Consejero Delegado Jefe de Equipo, Desarrollo de Software Registro #1: CEO Registro #2: Director Registro #3: Gerente	**Generalización:** Oficial de nivel C Gerente de TI **Intercambio:** Registro #1: Manager Registro #2: CEO Registro #3: Director
• Nombre de la empresa	**Generalización:** Generalizar el nombre de la empresa al sector industrial	Speedy Taxi Ltd	**Generalización:** Transporte y almacenamiento
	Intercambio: Cambie aleatoriamente los nombres de las empresas dentro del conjunto de datos.	Registro #1: Speedy Taxi Ltd Registro #2: Best Food Ltd Registro # 3: No. 1 Cold Wear Pte Ltd	**Intercambio:** Registro #1: Best Food Ltd Registro # 2: No. 1 Cold Wear Pte Ltd Registro # 3: Speedy Taxi Ltd

Identificadores indirectos	Técnica(s) de uso común	Ejemplo(s)	
		Antes	Después
• Dirección IP	**Enmascaramiento de caracteres:** Enmascara los dos últimos octetos[19] de direcciones IP IPv4 y los últimos 80 bits de direcciones IP IPv6. *Nota: El intercambio se puede aplicar además del enmascaramiento de caracteres.*	IPv4: 12.120.210.88 IPv6: 2001:0db8:85a3:0000:0000:8a2e:0370:7334	**Enmascaramiento de caracteres:** IPv4: 12.120.xxx.xxx IPv6: 2001:0db8:85a3:xxxx-:xxxx:xxxx:xxxx:xxxx:xxxx
• Número de matrícula del vehículo	**Enmascaramiento de caracteres:** Enmascarar los últimos cuatro caracteres del número de placa del vehículo. *Nota: El intercambio se puede aplicar además del enmascaramiento de caracteres.*	SMF1234A	**Enmascaramiento de caracteres:** SMF1xxxx
• Número de bastidor en el vehículo	**Enmascaramiento de caracteres:** Enmascara los últimos tres dígitos del número de bastidor. *Nota: El intercambio se puede aplicar además del enmascaramiento de caracteres.*	1234567890	**Enmascaramiento de caracteres:** 1234567xxx

19 Sugerimos enmascarar los dos últimos octetos independientemente de la clase de dirección de red (A/B/C) para evitar que la dirección enmascarada se identifique como perteneciente a una subred de clase B o C. También hace que sea más difícil agrupar a las personas que residen en la misma subred.

Identificadores indirectos	Técnica(s) de uso común	Ejemplo(s)	
		Antes	Después
• Ubicación del Sistema de Posicionamiento Global (GPS)	**Generalización:** Redondee las coordenadas GPS (en grados decimales) a los dos decimales más cercanos (equivalente a una precisión de 1,11 km) o a tres decimales (equivalente a una precisión de 111 m).	1.27434, 103.79967	**Generalización:** 1.274, 103.800 (grados decimales redondeados a tres decimales)
	Perturbación de datos: Agregue valores aleatorios entre 0.005 y -0.005 o entre 0.0005 y -0.0005.		**Perturbación de datos:** 1.27834, 103.79767
	Intercambio: Cambie aleatoriamente los valores de ubicación GPS dentro del conjunto de datos.	Registro #1: 1.27434, 103.79967 Registro #2: 1.26421, 103.80405 Registro #3: 1.26463, 103.82226	**Intercambio:** Registro #1: 1.26463, 103.82226 Registro #2: 1.27434, 103.79967 Registro #3: 1.26421, 103.80405

Atributos objetivo

Los atributos objetivo son información propietaria que es importante preservar para la utilidad de datos. Por lo tanto, para la mayoría de los casos de uso, las técnicas de anonimización no se aplican a los atributos objetivo. Sin embargo, para el caso de uso de datos sintéticos, dado que los datos a nivel de registro se utilizan normalmente en entornos de desarrollo y prueba que pueden no estar debidamente protegidos, se recomienda que se apliquen una o más técnicas de anonimización a los atributos objetivo para garantizar que no se produzca una nueva identificación en caso de brechas de datos.

Es importante comprobar y asegurarse de que, después de aplicar las técnicas de anonimización, ningún registro en el conjunto de datos sintéticos se parezca a ningún registro en el conjunto de datos original.

Atributos objetivo	Técnica(s) de uso común	Ejemplo(s)	
		Antes	Después
• Transacciones • Sueldo • Calificación crediticia • Póliza • Diagnóstico médico • Estado de vacunación	**Perturbación de datos:** Modifique aleatoriamente los datos numéricos (por ejemplo, sumando o restando valores aleatorios de los datos originales). La perturbación de datos no es posible para datos textuales alfanuméricos o no estructurados.	Valor de compra: €38.05 Salario: €1.200	**Perturbación de datos:** Valor de compra: €42 Salario: €1.700
	Intercambio: Cambie aleatoriamente los datos dentro del conjunto de datos. *Nota: El intercambio se puede aplicar además de la perturbación de datos.*	Estado de vacunación: Registro # 1: Vacunado Registro # 2: Primera dosis Registro # 3: No vacunado	**Intercambio:** Estado de vacunación: Registro # 1: Primera dosis Registro # 2: No vacunado Registro # 3: Vacunado

ANEXO C: K-ANONIMIDAD

k-anonimidad (y extensiones similares como l-diversity y t-closeness) es una medida utilizada para garantizar que no se haya superado el umbral de riesgo, como parte de la metodología de anonimización.

k-anonimidad no es la única medida disponible, ni está exenta de sus limitaciones, pero es relativamente bien entendida y fácil de aplicar. k-anonimidad puede no ser adecuado para todos los tipos de conjuntos de datos u otros casos de uso complejos. Otros enfoques y / o herramientas como el Algoritmo de Detección de Únicos Especiales (SUDA) y µ-Argus pueden ser más adecuados para evaluar el riesgo de grandes conjuntos de datos. Los métodos alternativos, como la privacidad diferencial[20], también han surgido en los últimos años.

k-anonimidad	
Descripción	El modelo de k-anonimidad se utiliza como guía antes de que se hayan aplicado técnicas de anonimización (por ejemplo, generalización), y para la verificación posterior también, para garantizar que los identificadores indirectos de cualquier registro sean compartidos por al menos k-1 otros registros. Esta es la protección clave proporcionada por k-anonimidad contra ataques de vinculación, porque los registros k (o al menos diferentes identificadores indirectos) son idénticos en sus atributos de identificación y, por lo tanto, crean una clase de equivalencia[21] con k miembros. Por lo tanto, no es posible vincular o señalar el registro de un individuo, ya que siempre hay k atributos idénticos. Un conjunto de datos anonimizado puede tener diferentes niveles de k-anonimidad para diferentes conjuntos de identificadores indirectos, pero para la protección de "máximo riesgo" contra la vinculación, la k más baja se utiliza como valor representativo para la comparación con el umbral.

20 La privacidad diferencial implica varios conceptos, incluida la respuesta a consultas en lugar de proporcionar el conjunto de datos anónimo, la adición de ruido aleatorio a la protección de registros individuales, la provisión de garantías matemáticas de que no se excede el "presupuesto de privacidad" predefinido, etc.

21 "Clase de equivalencia" se refiere a los registros de un conjunto de datos que comparten los mismos valores dentro de ciertos atributos, normalmente identificadores indirectos.

k-anonimidad

Cuando usarlo	k-anonimidad se utiliza para confirmar que las medidas de anonimización implementadas alcanzan el umbral deseado contra los ataques de enlace.
Cómo usarlo	Primero, decida un valor para k (que sea igual o superior al inverso del tamaño de la clase de equivalencia), que proporcione el k más bajo que se debe lograr entre todas las clases de equivalencia. En general, cuanto mayor es el valor de k, más difícil es para los interesados ser identificados; sin embargo, la utilidad puede volverse más baja a medida que k aumenta y es posible que se deban suprimir más registros. Después de que se hayan aplicado las técnicas de anonimización, verifique que cada registro tenga al menos k-1 otros registros con los mismos atributos abordados por la k-anonimización. Los registros en clases de equivalencia con menos de k registros deben considerarse para la supresión; alternativamente, el conjunto de datos se puede anonimizar aún más.
Otros consejos	• Además de la generalización y la supresión, también se pueden crear datos sintéticos para lograr el anonimato k. Estas técnicas (y otras) a veces se pueden usar en combinación, pero tenga en cuenta que el método específico elegido puede afectar la utilidad de datos. Considere las compensaciones entre abandonar los valores atípicos o insertar datos sintéticos. • k-anonimidad asume que cada registro se relaciona con un individuo diferente. Si el mismo individuo tiene múltiples registros (por ejemplo, visitando el hospital en varias ocasiones), entonces la k-anonimidad deberá ser más alta que los registros repetidos, de lo contrario los registros no solo pueden ser vinculables, sino que también pueden ser reidentificables, a pesar de cumplir aparentemente con las "clases de equivalencia k".

Ejemplo.
En este ejemplo, el conjunto de datos contiene información sobre las personas que toman taxis.

Se utiliza k = 5 (es decir, cada registro debe eventualmente compartir los mismos atributos con otros cuatro registros después de la anonimización).

Las siguientes técnicas de anonimización se utilizan en combinación. El nivel de granularidad es un enfoque para alcanzar el nivel k requerido.

Atributo	Anonimación técnica
Edad	Generalización (intervalos de 10 años)
Ocupación	Generalización (por ejemplo, tanto "administrador de bases de datos" como "programador" se generalizan a "TI")
Supresión de registros	Los registros que no cumplen con los criterios de k-anonimato de 5 después de que se hayan aplicado técnicas de anonimización (en este caso, generalización) se eliminan. Por ejemplo, el registro del banquero se elimina, ya que es el único valor de este tipo en "Ocupación".

Conjunto de datos antes de la anonimización				
Número de serie	**Edad**	**Género**	**Ocupación**	**Promedio de viajes por semana**
1	21	Femenino	Oficial Asistente de Protección de Datos	15
2	38	Masculino	Consultor Líder de TI	2
3	25	Femenino	Banquero	8
4	34	Masculino	Administrador de bases de datos	3
5	30	Femenino	Director de Privacidad	1
6	29	Femenino	Delegado Regional de Protección de Datos	5
7	38	Masculino	Programador	3
8	32	Masculino	Analista de TI	4
9	25	Femenino	Delegado Adjunto de Protección de Datos	2
10	23	Femenino	Gerente, Oficina de DPO	11
11	31	Masculino	Diseñador UX	0

El conjunto de datos se convierte en 5 anónimo después de la anonimización de la edad y la ocupación, y la supresión del valor atípico. (Las respectivas clases de equivalencia están resaltadas en diferentes colores):

Conjunto de datos después de la anonimización				
Número de serie	**Edad**	**Género**	**Ocupación**	**Promedio de viajes por semana**
1	De 21 a 30	Femenino	Delegado de Protección de Datos	15
2	De 31 a 40	Masculino	ESO	2
3	De 21 a 30	Femenino	Banquero	8
4	De 31 a 40	Masculino	ESO	3
5	De 21 a 30	Femenino	Delegado de Protección de Datos	1
6	De 21 a 30	Femenino	Delegado de Protección de Datos	5
7	De 31 a 40	Masculino	ESO	3
8	De 31 a 40	Masculino	ESO	4
9	De 21 a 30	Femenino	Delegado de Protección de Datos	2
10	De 21 a 30	Femenino	Delegado de Protección de Datos	11
11	De 31 a 40	Masculino	ESO	0

Nota: El número promedio de viajes por semana se toma aquí como ejemplo para un atributo objetivo, sin necesidad de anonimizar aún más este atributo

ANEXO D: EVALUACIÓN DEL RIESGO DE REIDENTIFICACIÓN

Hay varias formas de evaluar el riesgo de reidentificación, y estas pueden requerir cálculos bastante complejos que impliquen el cálculo de probabilidades.

En esta sección se describe un modelo simplificado, utilizando la k-anonimidad[22], y se hacen las siguientes suposiciones:

1. El modelo de lanzamiento no es público;

2. El atacante está motivado para vincular a un individuo al conjunto de datos anonimizado; y

3. El contenido de los datos anonimizados no se tiene en cuenta y el riesgo calculado es independiente del tipo de información que el atacante realmente tiene disponible.

En primer lugar, debe establecerse el umbral de riesgo. Este valor, que refleja una probabilidad, oscila entre 0 y 1. Refleja el nivel de riesgo que la organización está dispuesta a aceptar. Los principales factores que afectan al umbral de riesgo deben incluir el daño que podría causarse al interesado, así como el daño a la organización, si se produce una nueva identificación; pero también tiene en cuenta qué otros controles se han establecido para mitigar cualquier riesgo residual. Cuanto mayor sea el daño potencial, mayor debe ser el umbral de riesgo. No existen normas estrictas y rápidas sobre qué valores umbral de riesgo deben utilizarse; los siguientes son solo ejemplos.

Daño potencial	Umbral de riesgo
Bajo	0.2
Medio	0.1
Alto	0.01

Al calcular el riesgo de reidentificación, esta guía utiliza el "Prosecutor Risk", que asume que el atacante conoce a una persona específica en el conjunto de datos y está tratando de establecer qué registro en el conjunto de datos se refiere a esa persona.

La regla simple para calcular la probabilidad de reidentificación para un solo registro en un conjunto de datos, es tomar el inverso del tamaño de la clase de equivalencia del registro:

22 Los cálculos serían diferentes si se hicieran utilizando la privacidad diferencial o los controles tradicionales de revelación estadística, por ejemplo.

P (vincular individuo a un solo registro) = 1 / tamaño de clase de equivalencia del registro

Para calcular la probabilidad de reidentificación de cualquier registro en todo el conjunto de datos, dado que hay un intento de reidentificación, un enfoque conservador sería equipararlo a la probabilidad máxima de reidentificación entre todos los registros del conjunto de datos.

P (volver a identificar cualquier registro en el conjunto de datos) =
1 / Min. tamaño de clase de equivalencia en el conjunto de datos

Nota: Si el conjunto de datos ha sido k-anonimizado,
P (volver a identificar cualquier registro en el conjunto de datos) < = 1 / k

Podemos considerar tres escenarios de ataque de intrusos motivados:

1. El ataque deliberado a las personas con información privilegiada;
2. El reconocimiento involuntario por parte de un conocido; y
3. Una brecha de datos.

P (re-ID) = P (re-ID | intento de re-ID) x P (intento de re-ID)

donde P (re-ID | intento de re-ID) se refiere a la probabilidad de reidentificación exitosa, dado que hay un intento de reidentificación. Como se discutió anteriormente, podemos tomar P (re-ID | intento de re-ID) para ser (1 / Min. tamaño de clase de equivalencia en el conjunto de datos)

Por lo tanto, P (re-ID) = (1 / Min. tamaño de clase de equivalencia en el conjunto de datos) x P (intento de re-ID)

Para el escenario # 1, el ataque interno deliberado, asumimos que una parte que recibe el conjunto de datos intenta reidentificarlos. Para estimar P (intento de re-ID): la probabilidad de un intento de reidentificación, los factores a considerar incluyen el alcance de los controles de mitigación implementados, así como los motivos y recursos del atacante. En la tabla siguiente se presentan valores de ejemplo; una vez más, corresponde a la parte que anonimiza el conjunto de datos decidir sobre los valores adecuados para usar.

Escenario #1: el ataque interno deliberado P (intento de re-ID) = P(ataque interno)

		Motivación y recursos del atacante		
		Bajo	**Medio**	**Alto**
Alcance de los controles de mitigación	Alto	0.03	0.05	0.1
	Medio	0.2	0.25	0.3
	Bajo	0.4	0.5	0.6
	Ninguno	1.0	1.0	1.0

Los factores que afectan la motivación y los recursos del atacante pueden incluir:

1. Disposición a violar el contrato (suponiendo que el contrato que impida la reidentificación esté en su lugar)

2. Limitaciones financieras y de tiempo

3. Inclusión de personalidades de alto perfil (por ejemplo, celebridades) o datos confidenciales (por ejemplo, información crediticia) en el conjunto de datos

4. Facilidad de acceso a datos o información "enlazables", ya sean de acceso público o de propiedad privada, que pueden permitir la reidentificación del conjunto de datos anonimizado

Los factores que afectan el alcance de los controles de mitigación incluyen:

1. Estructuras organizativas

2. Controles administrativos/legales (por ejemplo, contratos)

3. Controles técnicos y de procesos

Para el escenario # 2, el reconocimiento involuntario por parte de un conocido, asumimos que una parte que recibe el conjunto de datos vuelve a identificar inadvertidamente a un sujeto de datos mientras examina el conjunto de datos. Esto es posible porque la parte tiene algún conocimiento adicional sobre el interesado debido a su relación (por ejemplo, amigo, vecino, pariente, colega, etc.). Para estimar P (intento de re-ID): la probabilidad de un intento de reidentificación, el factor principal a considerar es la probabilidad de que el destinatario de los datos conozca a alguien en el conjunto de datos.

Escenario #2: reconocimiento involuntario por parte de un conocido P (intento de re-ID) = P (destinatario de datos que conoce a una persona en el conjunto de datos)

Para el escenario # 3, la probabilidad de que ocurra una brecha de datos en el sistema TIC del receptor de datos se puede estimar en función de las estadísticas disponibles sobre la prevalencia de brechas de datos en la industria del receptor de datos. Esto se basa en la suposición de que los atacantes que obtuvieron el conjunto de datos intentarán la reidentificación.

Escenario #3: una brecha de datos P (intento de re-ID) = P (brechas de datos en la industria del destinatario de datos)

La probabilidad más alta entre los tres escenarios debe usarse como P (intento de re-ID).

P (intento de re-ID) = Max (P (ataque interno),
P (destinatario de datos que conoce a una persona dentro del conjunto de datos),
P (brecha de datos en la industria del destinatario de datos))

Recopilando todo,

P (re-ID) = (1 / Min. tamaño de clase de equivalencia en el conjunto de datos)
x P (intento de re-ID) = (1 / k)
x P (intento de re-ID) para el conjunto de datos k-anonimizado

donde P (intento de re-ID) = Max (P (ataque interno),
P (destinatario de datos que conoce a una persona en el conjunto de datos),
P (brechas de datos en la industria del destinatario de datos))

ANEXO E: HERRAMIENTAS DE ANONIMIZACIÓN

La siguiente es una lista de algunas herramientas de anonimización comerciales o de código abierto.

Herramienta	Descripción	URL
Amnesia	La herramienta de anonimización de Amnesia es un software utilizado localmente para anonimizar datos personales y confidenciales. Actualmente admite garantías de k-anonimidad y km-anonimidad.	https://amnesia.openaire.eu/
Arcad DOT- Anonymizer	Anonymizer es una herramienta que mantiene la confidencialidad de los datos de prueba ocultando información personal. Funciona anonimizando los datos personales conservando su formato y tipo.	https://www.arcadsoftware.com/dot/data-masking/dot-anonymizer/
ARGUS	ARGUS significa "Anti Re-identification General Utility System". La herramienta utiliza una amplia gama de diferentes métodos de anonimización estadística, como la recodificación global (agrupación de categorías), la supresión local, la aleatorización, la adición de ruido, la micro agregación, la codificación superior e inferior. También se puede utilizar para generar datos sintéticos.	https://research.cbs.nl/casc/mu.htm
ARX	ARX es un software de código abierto para anonimizar datos personales confidenciales.	https://arx.deidentifier.org/

Herramienta	Descripción	URL
Eclipse	Eclipse es un conjunto de herramientas de Privacy Analytics que facilita la anonimización de los datos de salud.	https://privacy-analytics.com/health-data-privacy/health-data-software/
sdcMicro	sdcMicro se utiliza para generar microdatos anonimizados, como archivos de uso público y científico. Admite diferentes métodos de estimación de riesgos.	https://cran.r-project.org/web/packages/sdcMicro/index.html
UTD Anonymisation Toolbox	UT Dallas Data Security and Privacy Lab compiló varias técnicas de anonimización en una caja de herramientas para uso público.	http://cs.utdallas.edu/dspl/cgi-bin/toolbox/index.php?go=home

AGRADECIMIENTOS

El PDPC y Infocomm Media Development Authority (IMDA) expresan su sincero agradecimiento a las siguientes organizaciones por sus valiosos comentarios en el desarrollo de esta publicación.

- AsiaDPO
- BetterData Pte Ltd
- ISACA (Singapore Chapter) — SIG de protección de datos
- Law Society of Singapore — Comité de Ciberseguridad y Protección de Datos (CSDPC)
- Ministerio de Salud (MOH)
- Replica Analytics
- Privitar Ltd
- SGTech
- Federación Empresarial de Singapur (SBF) — Comité de Digitalización
- Singapore Corporate Counsel Association (SCCA) — Capítulo de Protección de Datos, Privacidad y Ciberseguridad (DPPC)
- Departamento de Estadística de Singapur (DOS)
- Smart Nation and Digital Government Group (SNDGO)

En esta guía se hizo referencia a las siguientes guías.

•UKAN. The Anonymisation Decision Making Framework 2nd Edition: European Practitioners' Guide, por Mark Elliot, Elaine Mackey y Kieron O'Hara, 2020.
•CSIRO y OAIC. The De-Identification Decision-Making Framework, por Christine M O'Keefe, Stephanie Otorepec, Mark Elliot, Elaine Mackey y Kieron O'Hara, 18 de septiembre de 2017.
•IPC. Directrices de desidentificación para datos estructurados, junio de 2016, **https://www.ipc.on.ca/wp-content/uploads/2016/08/Deidentification-Guidelines-for-Structured-Data.pdf.**
•El Emam, K. Guide to the De-Identification of Personal Health Information, CRC Press, 2013.
•Artículo 29 Grupo de Trabajo sobre Protección de Datos (Comisión Europea). "Dictamen 05/2014 sobre técnicas de anonimización". 10 de abril de 2014, **http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf.**
•NIST. NISTIR 8053: Desidentificación de información personal, por S L Garfinkel, octubre de 2015, **http://nvlpubs.nist.gov/nistpubs/ir/2015/NIST.IR.8053.pdf.**

www.aepd.es

@aepd_es