ARCHIVO GENERAL DE LA NACIÓN COLOMBIA

GUÍA DE ANONIMIZACIÓN DE DATOS ESTRUCTURADOS

Conceptos generales y propuesta metodológica

La cultura es de todos

Mincultura

ARCHIVO GENERAL DE LA NACIÓN
JORGE PALACIOS PRECIADO - COLOMBIA
Establecimiento público
adscrito al Ministerio de Cultura

Archivo General de la Nación
Enrique Serrano López
Director General

AUTORES
Erika Lucia Rangel
Archivo General de la Nación de Colombia

Diana Ramírez Roa
Departamento Nacional de planeación

Claudia Rodriguez
Ministerio de Tecnologías de la Información y las Comunicaciones

COLABORADORES
Superintendencia de Industria y Comercio
Departamento Administrativo de la Función Pública
Departamento Administrativo Nacional
Departamento Administrativo Nacional de Estadística -
Agustín Jimenez Ospina

Revisión de textos, diseño y diagramación
María Angélica Osorio
Catalina Lozano Ortega

Imágenes
Fotos de Tecnología www.freepik.es

ISBN
En trámite

Archivo General de la Nación de Colombia
Carrera 6 No. 6-91
Teléfono: 328 2888 Fax: 337 2019
E-mail: contacto@archivogeneral.gov.co
Página web: www.archivogeneral.gov.co
Bogotá D.C., Colombia - 2020

Contenido

Índice de ilustraciones

Índice de tablas

1. Introducción

Volver al índice

1.1 Introducción

En el marco de la Política Nacional de Explotación de Datos CONPES 3920 de 2018 (Consejo Nacional de Política Económica y Social - CONPES, 2018) y la Política de Transformación Digital e Inteligencia Artificial CONPES 3975 de 2019, se **reconoce el potencial que tienen los datos como activo para la generación de valor social y económico**. Por tal motivo, uno de sus objetivos es que las entidades masifiquen la disponibilidad de datos públicos digitales accesibles, usables y de calidad. En ese sentido, se identifica la necesidad de fortalecer la apertura de datos para su reutilización, salvaguardando los principios reconocidos por la legislación en materia de transparencia y protección de datos personales.

La aparición de nuevas tecnologías de datos como Big Data, Inteligencia Artificial e Internet de las cosas, incorpora nuevas oportunidades para la creación de valor público, pero también genera grandes desafíos en materia de protección de datos personales.

En Colombia, la Ley General de protección de datos personales- Ley 1581 de 2012 (Congreso de la República, 2012) establece los principios aplicables a las actividades de tratamiento de datos personales para garantizar el derecho fundamental de Habeas Data de las personas.

Dentro de ese listado de principios previsto en el artículo cuatro de la citada Ley se incluyen los principios de acceso y circulación restringida y seguridad que señalan que los datos personales, a excepción de la información pública, no pueden estar disponibles en internet u otros medios de divulgación o comunicación masiva, salvo que el acceso sea técnicamente controlable y que se debe contar con las medidas técnicas, humanas y administrativas para evitar su adulteración, pérdida, uso y acceso no autorizado de la información.

El tratamiento de grandes volúmenes de datos contribuye a la generación de valor social y económico siempre que se respeten los derechos de las personas frente a su privacidad y protección de datos personales. En este contexto, la anonimización se incorpora como un proceso para explotar los datos eliminando la posibilidad de identificación de las personas.

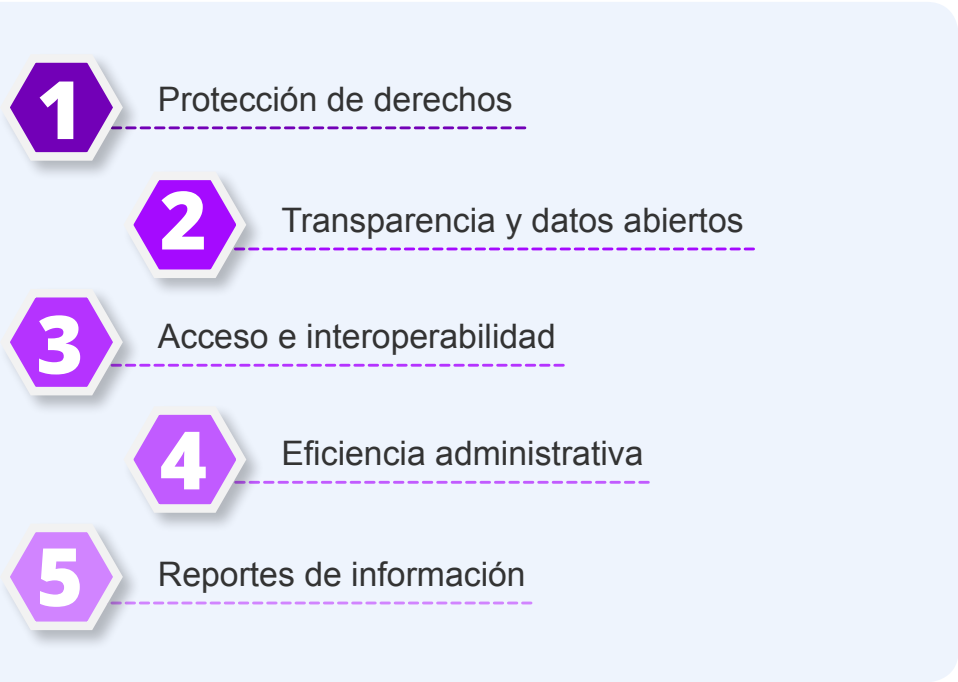
> La finalidad del proceso de anonimización es evitar la identificación de las personas y reducir su probabilidad de reidentificación sin afectar la veracidad de los resultados y la utilidad de los datos que han sido tratados. Este procedimiento es especialmente relevante en los entornos que surgen con la evolución tecnológica y fenómenos como Big Data u Open Data, los cuales aumentan la probabilidad de reidentificación de las personas.

En este contexto, la anonimización surge como un instrumento para minimizar los riesgos que se presentan en el tratamiento masivo de los datos de carácter personal. Por consiguiente, este proceso permite adoptar procedimientos acordes a las normas y en beneficio de las empresas y el comercio, sin contravenir el ámbito privado de sus titulares. Esto implica la adopción de procesos que permitan evitar la identificación de las personas y tratar la información que pueda generar cualquier afectación o violación a los derechos de los titulares de esta.

En la elaboración el documento **CONPES 3920**, se identificó que solo el 31,8% de las entidades del orden nacional ha aplicado alguna técnica de anonimización de datos personales, lo que permite concluir que es necesario fortalecer la cultura de anonimización, así como fortalecer y disponer de herramientas técnicas y metodológicas que faciliten la apertura de los datos, en el marco del régimen de protección de datos personales.

De acuerdo con lo anterior, el **Archivo General de la Nación**, en articulación con el **Ministerio de Tecnologías de la Información y las Comunicaciones**, la **Superintendencia de Industria y Comercio**, el **Departamento Administrativo de la Función Pública**, el **Departamento Nacional de Planeación** y el **Departamento Administrativo Nacional de Estadística**, desarrolló esta guía con el fin de precisar conceptos y proporcionar una orientación metodológica para realizar procesos de anonimización de datos personales e información producidos o gestionados por entidades públicas y privadas con funciones públicas.

Esta guía se enmarca en las disposiciones jurídicas nacionales e internacionales en materia de privacidad, protección de datos personales, acceso a la información pública y transparencia. Su aplicación le corresponde a las entidades públicas y a las empresas privadas que desarrollan funciones públicas y que tienen procesos de tratamiento de datos personales. En ese sentido, la guía busca brindar elementos metodológicos y técnicos para que las entidades garanticen la protección de cualquier información producida, gestionada o recolectada que contenga datos personales bajo las siguientes premisas:



Este documento está dividido en **5 capítulos** siendo la introducción el primero de ellos. En la segunda parte se presenta un marco conceptual de la anonimización de datos personales incorporando el marco jurídico en Colombia aplicable al proceso técnico. Posteriormente se expone la metodología propuesta para la anonimización de datos personales. En la última sección se describen recomendaciones y por último se incluyen las referencias bibliográficas.

1.2 Objetivo de la guía

01 Precisar conceptos y proporcionar una orientación metodológica a las entidades públicas y privadas con funciones públicas, respecto a la anonimización de datos personales disponibles en bases de datos estructurados.

02 Contribuir a que las entidades públicas o privadas con funciones públicas den cumplimiento a la normatividad existente en materia de protección de datos personales y simultáneamente logren el aprovechamiento de datos como activo, garantizando el derecho de acceso a la información pública.

03 Proporcionar orientación sobre las técnicas de anonimización y su uso adecuado según las características de los conjuntos de datos que requieran ser anonimizados, a las entidades públicas que requieran procesar datos para una serie de propósitos incluido el análisis de Big Data.

1.3 Alcance de la guía

La orientación metodológica y técnicas de anonimización de datos personales que se presentan en esta guía están orientadas para datos estructurados. No incluye lineamientos para la anonimización de datos no estructurados como imágenes, video, audio o demás archivos multimedia. Se incorporará una segunda guía sobre lineamientos para la anonimización de datos personales en documentos físicos y datos en documentos textuales de archivo.

1.4 Público Objetivo

La guía está dirigida a las entidades y organizaciones que realicen tratamiento de datos personales, es decir que gestionen, almacenen, administren, obtengan, produzcan, procesen, custodien y publiquen información independientemente de su soporte o medio y que deban dar cumplimiento a la normatividad de protección de datos personales.

2. Anonimización de datos personales

Volver al índice

2.1 ¿Qué son los datos personales?

Según la Ley 1581 de 2012, un dato personal se define como cualquier información que pueda asociarse a una o varias personas naturales determinadas o determinables. Una persona o individuo puede ser identificado directa o indirectamente a través de su nombre, número de identificación, datos de ubicación, información laboral, entre otros.

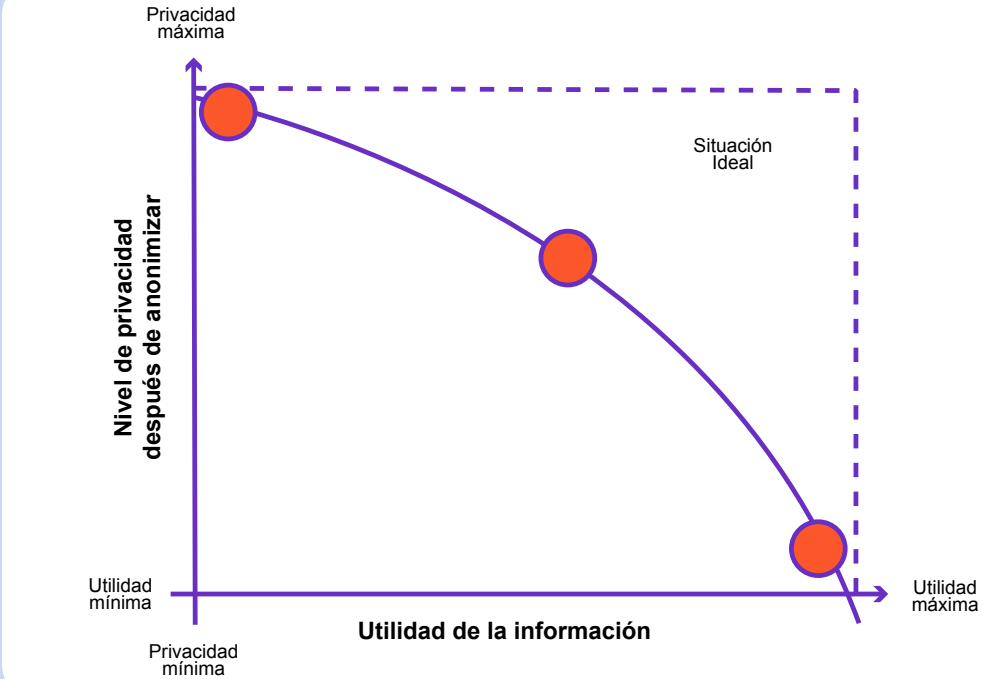
2.2 ¿Qué es anonimizar datos?

La anonimización es el proceso mediante el cual se condiciona un conjunto de datos de modo que no se pueda identificar a una persona, pero pueda ser utilizada para realizar análisis técnico y científico válido sobre ese conjunto de datos (MIT, 2007). Para el cumplimiento de los estándares de anonimización, los datos deben ser despojados de elementos suficientes para que el titular de los datos ya no pueda ser identificado, y por lo tanto estos datos deben procesarse para que no sea posible identificar a una persona mediante el uso de todos los medios razonables para ser utilizados por cualquier otra persona (Data Protection Commission, 2019).

El proceso de anonimización de datos personales requiere una adecuada comprensión del propósito final de la utilización de la información, así como de su nivel de utilidad, teniendo en cuenta que independientemente de las técnicas empleadas, una vez realizado el proceso de anonimización se reduce la información original del conjunto de datos. Por tal motivo es importante que la entidad administradora de los datos decida el costo de oportunidad entre la utilidad que se busca obtener a partir de los datos y el nivel de riesgo de reidentificación (PDPC, 2018).

Como se observa en la Ilustración 1, las entidades y empresas se enfrentan a estas dos variables al momento de anonimizar datos para su tratamiento y publicación: riesgo de reidentificación y utilidad de la información.

Ilustración 1. Riesgo de reidentificación vs utilidad de la información



Fuente: Elaboración propia tomada de Luk Arbuckle, Privacy Analytics

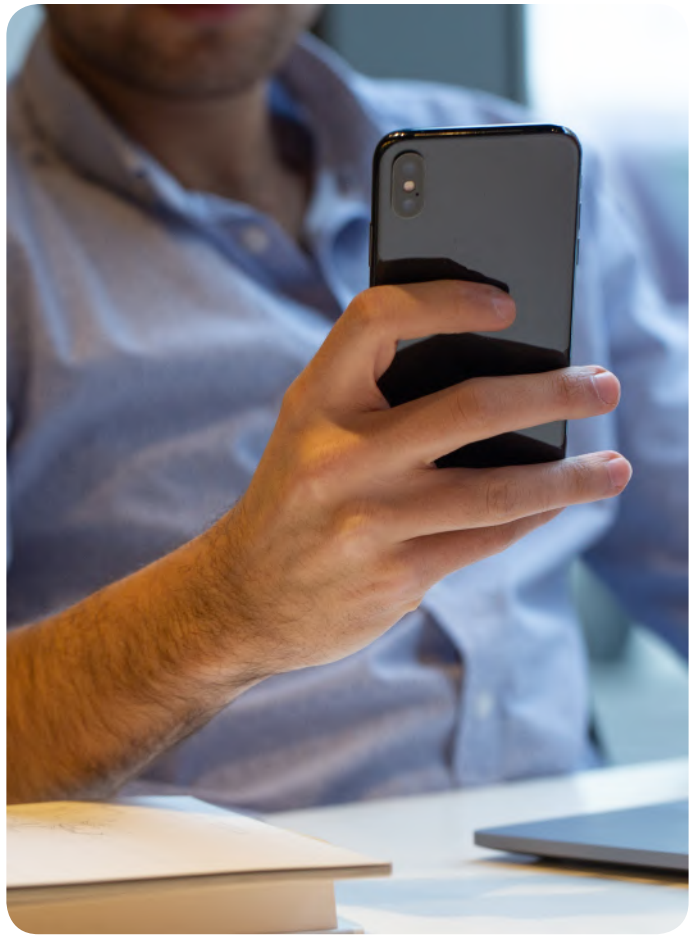
La Ilustración 1, permite observar la relación entre el nivel de privacidad después de anonimizar y la utilidad de la información. A mayor nivel de información no anonimizada mayor es el riesgo de reidentificación, aunque mayor es su utilidad; por tal motivo, se debe buscar en los procesos de anonimización, máximos riesgos tolerables y el nivel de privacidad de la información logrando un equilibrio para obtener utilidad de los datos después de anonimizados. Una de las consideraciones más importantes que permitan maximizar la utilidad de la información, controlando a su vez la posibilidad de riesgo de identificación a tener en cuenta en el proceso es realizar una prueba de "identificabilidad" a la técnica de anonimización.

2.3 ¿Cúando se considera que los datos son anonimizados?

Los datos se encuentran anonimizados cuando los sujetos dueños de los datos ya no pueden ser identificables, inclusive con la ayuda de los datos originales. Una vez los datos son anonimizados, estos se pueden usar, reutilizar y divulgar sin violar el derecho a la protección de datos de los titulares de la información. Con el desarrollo del Big Data, cada vez hay menos probabilidad de garantizar que una persona no pueda ser identificable a partir de un conjunto de datos que haya sido anteriormente sometido a técnicas de anonimización. Sin embargo, una técnica de anonimización efectiva reducirá el riesgo y la probabilidad de reidentificación y será capaz de prevenir la diferenciación de sujetos y la vinculación de registros asociados a este.

2.4 ¿Por qué se debe anonimizar?

- Para impedir que, a partir de un dato o de una combinación de datos de una misma fuente o de diferentes fuentes de datos, se logre identificar sujetos individuales ya sean individuos, empresas o establecimientos, u otro tipo de unidades de observación. (DANE, 2018).
- Para proteger los derechos de los titulares de los datos e información y reducir o eliminar definitivamente el riesgo de reidentificación, cuando se publica o coloca a disposición información para su consulta de manera abierta.
- Para evitar la identificación directa y la identificación indirecta. La identificación directa proviene de información como el nombre, la dirección, los números de teléfono y la identificación indirecta se obtiene del cruce de datos y otras fuentes de información.
- Para facilitar la divulgación, publicación e intercambio de datos, sin vulnerar los derechos a la protección de datos, de manera que no se puedan identificar directa o indirectamente las personas asociadas a los datos personales o características.
- Para publicar de manera segura datos abiertos protegiendo la privacidad de las personas. Los datos abiertos incrementan la transparencia del gobierno, y permiten que cualquier ciudadano pueda hacer uso y reutilización de conjuntos de datos con diferentes propósitos entre los que se encuentra la innovación y desarrollo de productos y servicios.
- Para cuándo se intercambia información con otras entidades y este intercambio requiere la inclusión de datos personales.



2.5 La protección de datos personales en el entorno de la transformación digital

La disponibilidad de datos masivos a través de la digitalización y el despliegue de servicios mediante el Internet ha posicionado a los datos como un activo fundamental para la generación de valor social y económico, de tal forma que se han convertido en uno de los activos más preciados en la actualidad para compañías y entidades públicas, pues facilitan la toma de decisiones en materia de productividad y eficiencia.

Los datos generados en el marco de las nuevas dinámicas tecnológicas se caracterizan por producirse en grandes volúmenes, provenir de distintas fuentes y formatos de información y producirse a grandes velocidades. El conjunto de tecnologías que permiten tratar grandes volúmenes de datos generados a grandes velocidades y por múltiples fuentes de información permiten transformar los datos en información y de esta forma aportar soluciones y oportunidades a asuntos de carácter privado y/o de política pública. Especialmente a partir del tratamiento de datos no estructurados de los que anteriormente no se podía extraer valor agregado, ni disponer un análisis sistemático (como por ejemplo los datos no texto como fotografías, videos, audios).

Sin embargo, a pesar de los beneficios que ofrece el Big Data, surgen desafíos relacionados con el tratamiento de datos personales

y la vulneración de los derechos de la protección de datos personales. Específicamente el principal reto se genera a partir de la alta probabilidad de riesgo de re- identificación de las personas, dada la facilidad para relacionar diferentes conjuntos de datos. Por ejemplo, una base de datos previamente anonimizada puede combinarse con otros datos de forma que se logre reidentificar a los individuos.

Por tal motivo, surge la necesidad de garantizar el uso de grandes volúmenes de datos, en el marco de técnicas y métodos de anonimización que permitan garantizar la reducción del riesgo y el cumplimiento de la normatividad vigente en términos de protección de datos personales.

2.6 ¿Cúales son los principios de la anonimización?

En un proceso de anonimización se requiere que las entidades definan un protocolo y unos procedimientos específicos para este fin. La anonimización debe convertirse en una práctica constante que cumpla con las medidas de protección de los datos desde los principios de privacidad por diseño y la privacidad por defecto, los cuales son medida s proactivas para responder al Principio de Responsabilidad Demostrada.

De acuerdo con el decreto 1074 de 2015, los responsables de la recolección de datos personales deberán proveer una descripción de los procedimientos usados para la recolección, almacenamiento, uso, circulación y supresión de la información, y describir la finalidad para la cual se recolectó la información. Así mismo, en la sección 6 se establece el principio de responsabilidad demostrada por el cual los responsables de tratamientos de datos personales deben tener la capacidad de demostrar que han implementado medidas efectivas y pertinentes para cumplir con las obligaciones establecidas en la Ley 1581 de 2012.

A continuación, se describe el principio de privacidad por diseño y por defecto:

A Principio de privacidad por diseño

Este principio se aborda desde un enfoque organizacional, en el cual las medidas de privacidad no deben estar únicamente en función del cumplimiento de la normatividad vigente, sino deben responder a una estrategia predeterminada dentro de una entidad y organización para la protección de datos personales. En este sentido, es una medida proactiva en lugar de reactiva, de manera que se anticipa la pérdida de privacidad de la información antes de que llegue a suceder.

Las entidades deben implementar las garantías al adecuado tratamiento de los datos personales durante todo el ciclo de ejecución de un proyecto y durante el ciclo de vida de los datos. Por tanto, se requiere la definición de lineamientos, procesos, procedimientos y la implementación de mecanismos adecuados para la protección de datos desde el mismo diseño de bases de datos, sistemas de información, formularios entre otros.

Las medidas de protección de privacidad se deben garantizar desde el inicio del ciclo de vida de los datos, desde que estos no están anonimizados. Una vez se inicia la explotación de la información anonimizada se debe continuar implementando medidas para la protección de los datos personales.

Acogiendo este principio es importante hacer una clasificación inicial de los datos, identificar las técnicas de anonimización, identificar a los responsables del tratamiento de los datos y garantizar que ellos estén en condiciones de garantizar la protección de estos. Cabe resaltar, que la obligación de que se adopte el principio de privacidad por diseño recae principalmente en el responsable del tratamiento de los datos personales, dado que esta acción no se puede delegar y en colaboración con los arquitectos de información y sistemas de información de las entidades o los que hagan sus veces.

Las medidas de seguridad para la implementación de este principio deben considerar los siguientes factores: **1) Los niveles del riesgo del tratamiento para los derechos de los titulares de los datos, 2) la naturaleza de los datos, 3) las consecuencias que se originen de una vulneración para los titulares de los datos, evaluando los daños y perjuicios causados, 4) el número de titulares de los datos, 5) el tamaño de la organización, 6) la fiabilidad de la técnica aplicada al tratamiento de datos, 7) el alcance y objetivo del tratamiento de datos. (SIC, 2019)**

Un ejemplo de protección datos por diseño, es la implementación de técnicas de seudonimización desde el desarrollo y programación de las aplicaciones de *software* (reemplazar los datos de identificación personal con identificadores artificiales), las cuales generan conjuntos de datos seudo anonimizados antes de ser publicados o consultados según roles y encriptación o cifrado (codificación de mensajes para que solo aquellas personas autorizadas puedan leer la información). (European Commission, 2019).

B Principio de privacidad por defecto

Este principio parte de atender al principio de proporcionalidad y necesidad de los datos. Es decir, considerando la recolección de datos estrictamente necesarios para la finalidad que se quiere conseguir, y evitar recabar datos de manera indiscriminada, lo cual aumenta la posibilidad de comprometer la privacidad de las personas.

Por defecto, las entidades deben garantizar que los datos personales se traten con la mayor protección a la intimidad, y deben procesar únicamente los datos personales necesarios para el propósito específico del tratamiento.

Un ejemplo de este principio es cuando un navegador tiene la opción de aceptar o rechazar cookies, en este caso la opción por defecto debe ser la segunda la cual garantiza la mayor protección de la privacidad, o por ejemplo cuando una plataforma de redes sociales limita desde el inicio y de forma predeterminada la accesibilidad al perfil de los usuarios, de forma que este perfil no sea accesible por defecto a un número indefinido de personas.

Este principio abarca: **i) La cantidad de datos recogidos, ii) la extensión de su tratamiento, iii) el plazo de su conservación, y iv) la accesibilidad.**

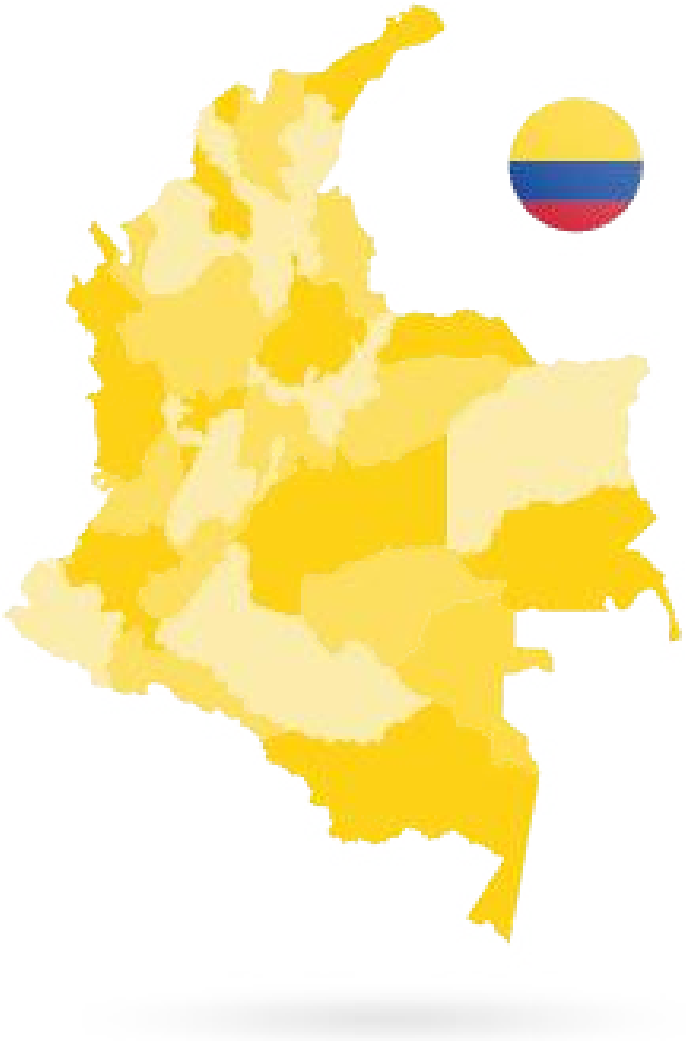
2.7 Marco Jurídico

La anonimización de datos es un proceso que permite mitigar los riesgos asociados al tratamiento de datos personales, lo cual reduce la probabilidad de vulnerar los derechos a la protección de datos de las personas. A continuación, se referencia el marco normativo aplicable a la protección de datos personales a nivel internacional y para Colombia.

2.7.1 Regulación internacional en materia de protección de datos e información

A nivel internacional dos de las legislaciones que han desarrollado en mayor medida el derecho a la protección de datos personales se basan en principios diferentes. La legislación europea está basada en el derecho fundamental a la privacidad y el Habeas Data, mientras que la legislación de Estado Unidos está derivada del principio de responsabilidad. En términos generales, la protección de datos en la Unión Europea está conformado por amplias garantías respaldadas por la constitución como derechos fundamentales integrales, y sus principios se aplican independientemente del contexto. Por el contrario, en Estados Unidos, las garantías de protección de los datos son casuísticas, específicas del contexto y de los sectores, varían según los instrumentos existentes y son mucho menos completas (European Parliament, 2015).

Por ejemplo, en el caso de la Unión Europea la compartición de datos entre agencias compromete los derechos fundamentales de las personas y requieren de una justificación concreta, mientras que en Estados Unidos el intercambio de datos sin mayor restricción es más la regla que la excepción (European Parliament, 2015). Otra importante distinción entre ambos marcos se determina por el alcance de la ley de protección de datos personales. En la Unión Europea la posible vulneración de derechos fundamentales a partir del tratamiento de datos brinda la posibilidad inmediata para que el individuo inicie un proceso judicial, mientras que, en Estados Unidos, la recopilación masiva de datos no conduce inmediatamente a otorgar una acción jurídica al sujeto titular de sus datos. ***(Ver Anexo 2)***

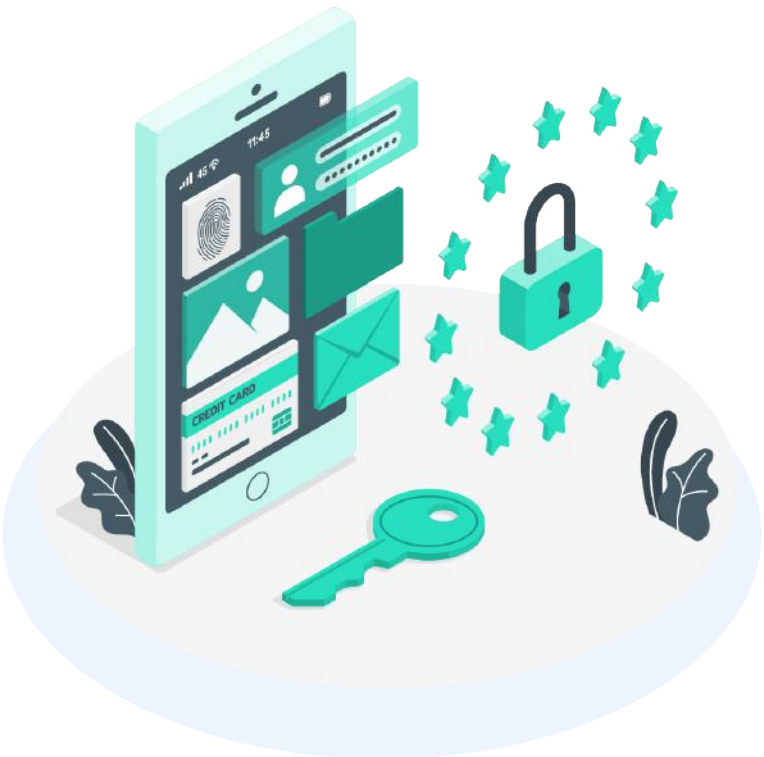


2.7.2 Marco Regulatorio en Colombia

En la Constitución Política de Colombia de 1991, específicamente en el artículo 15, se establece el derecho que tienen todas las personas a conservar su intimidad, mantener su buen nombre y a la protección y garantía del Habeas Data. En relación con la protección de datos personales, el régimen normativo colombiano dispone de la Ley 1266 de 2008 y la Ley 1581 de 2012.

A partir de la Ley 1266 de 2008, se crearon en el país las disposiciones del Habeas Data y la regulación del manejo de información de datos personales, relacionada especialmente con la información financiera, comercial y crediticia. Esta Ley tiene aplicabilidad en todos los datos de información personal administrados por las entidades públicas y privadas y les otorga a los ciudadanos a partir del articulo 14 y 15, el derecho a la protección de datos personales a partir de los siguientes procedimientos:

- Consulta de sus datos.
- Solicitar la actualización, corrección y eliminación de la base de datos cuando no existe ninguna disposición legal para mantenerla.



Esta ley definió la **tipología de los datos de carácter personal** (Superintendencia de Industria y Comercio, 2014):

* **Dato íntimo o privado**

  Es el dato que por su naturaleza íntima o reservada solo es relevante para el Titular.

* **Dato semiprivado**

  Es semiprivado el dato que no tiene naturaleza íntima, reservada, ni pública y cuyo conocimiento o divulgación puede interesar no sólo a su Titular sino a cierto sector o grupo de personas o a la sociedad en general. Un ejemplo de este tipo de datos son las historias crediticias de las personas.

* **Dato público**

  Es el dato calificado como tal según los mandatos de la Ley o de la Constitución Política y todos aquellos que no sean semiprivados o privados. Son públicos los datos que están relacionados con un interés general. Entre los datos públicos se encuentran los contenidos en documentos públicos, sentencias judiciales debidamente ejecutoriadas que no estén sometidos a reserva y los relativos al estado civil de las personas.

Adicionalmente, la Ley 1581 de 2012 que establece los **principios rectores que rigen el tratamiento de datos personales bajo custodia** de cualquier persona jurídica, pública o privada, estableció las siguientes categorías de datos personales:

* **Dato sensibles**

Aquellos que afectan la intimidad del titular o cuyo uso indebido puede generar su discriminación, tales como aquellos que revelen el origen racial o étnico, la orientación política, las convicciones religiosas o filosóficas, la pertenencia a sindicatos, organizaciones sociales, de derechos humanos o que promueva intereses de cualquier partido político o que garanticen los derechos y garantías de partidos políticos de oposición, así como los datos relativos a la salud, a la vida sexual y los datos biométricos.

* **Datos personales de los niños, niñas y adolescentes**

Se prohíbe el tratamiento de los datos personales de los niños, niñas y adolescentes, salvo aquellos que por su naturaleza son públicos. Sin embargo, la Corte Constitucional precisó que independientemente de la naturaleza del dato, se puede realizar el tratamiento de estos siempre y cuando el fin que se persiga con dicho tratamiento responda al interés superior de los niños, niñas y adolescentes y se asegure sin excepción alguna el respeto a sus derechos prevalentes.

La Ley 1581 de 2012 también define los siguientes roles y definiciones:

* **Tratamiento de datos**

  Cualquier operación o conjunto de operaciones sobre datos personales, tales como la recolección, almacenamiento, uso, circulación o supresión.

* **Responsable de Tratamiento**

  “Persona natural o jurídica, pública o privada, que por sí misma o en asocio con otros, decida sobre la base de datos y/o el Tratamiento de los datos”.

* **Encargado del Tratamiento**

  “Persona natural o jurídica, pública o privada, que por sí misma o en asocio con otros, realice el Tratamiento de datos personales por cuenta del responsable del Tratamiento”.

* **Titular**

  Persona natural cuyos datos sean objeto de tratamiento.

En el año 2012, a partir de la Ley 1581 de 2012, se definió la regulación vigente para la protección del derecho fundamental que tienen las personas naturales para autorizar la información personal que se encuentra almacenada en las bases de datos, su posterior actualización y rectificación. Los principios y disposiciones de esta Ley aplican a todos los datos personales registrados en bases de datos que los hagan susceptibles a ser tratados por entidades privadas y públicas.

Ámbito de aplicación de la Ley 1581 de 2012:

Esta ley se aplica a las bases de datos de datos personales de personas naturales. Se exceptúan las siguientes bases de datos:

a. **Bases de datos o archivos mantenidos en un ámbito exclusivamente personal o doméstico.** Según el artículo 2 del Decreto 1377 de 2013, el ámbito personal o doméstico comprende aquellas actividades que se inscriben en el marco de la vida privada o familiar de las personas naturales. Cuando estas bases de datos o archivos vayan a ser suministrados a terceros se deberá, de manera previa, informar al Titular y solicitar su autorización. En este caso los responsables y encargados de las bases de datos y archivos quedarán sujetos a las disposiciones contenidas en la presente ley.

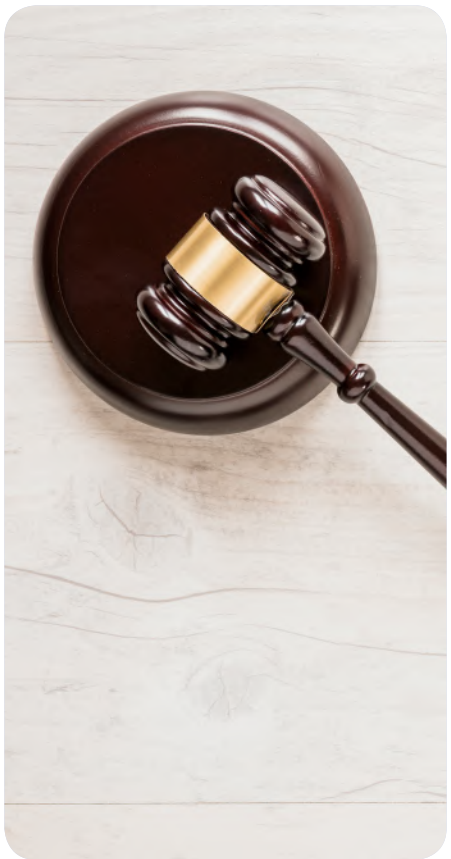
b. A las bases de datos y archivos que tengan por finalidad la seguridad y defensa nacional, así como la prevención, detección, monitoreo y control del lavado de activos y el financiamiento del terrorismo.

c. A las bases de datos que tengan como fin y contengan información de inteligencia y contrainteligencia.

d. A las bases de datos y archivos de información periodística y otros contenidos editoriales.

e. A las bases de datos y archivos regulados por la Ley 1266 de 2008 (datos financieros y económicos).

2.7.3 Principios relativos al tratamiento de datos



A partir de la Ley 1266 de 2008 y la Ley 1581 de 2012, se estipularon los siguientes principios aplicados a la administración y tratamiento de datos:

Principio de Legalidad

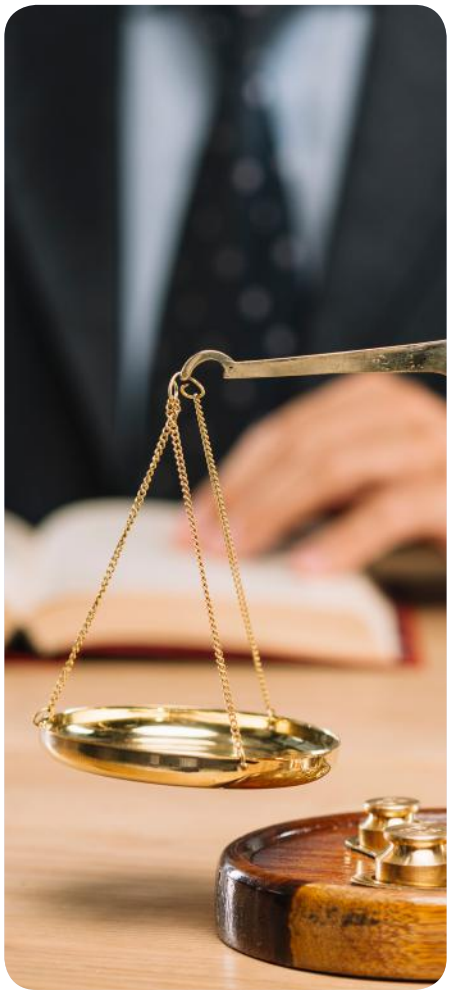
El tratamiento de los datos es una actividad reglada que está sujeta a lo establecido en la ley.

Principio de veracidad o calidad de los registros o datos

La información que está incluida en los bancos de datos deber ser veraz, completa, exacta, actualizada y comprobable. Se prohíbe la incorporación o divulgar datos parciales, incompletos, fraccionados o que induzcan a error.

Principio de transparencia

El tratamiento de datos debe garantizarse el derecho del Titular a obtener del encargado del tratamiento, en cualquier momento y sin ninguna restricción, información sobre la existencia de datos que le conciernan.



Principio de finalidad

La utilización de los datos debe sujetarse a una finalidad legítima de acuerdo con la Constitución y la ley. La finalidad de la utilización de los datos debe ser informada al titular de la información previa o concomitantemente con el otorgamiento de la autorización, cuando ella sea necesaria o en general siempre que el titular solicite información al respecto.

Principio de libertad

El Tratamiento solo puede ejercerse con el consentimiento, previo, expreso e informado del Titular. Los datos personales no podrán ser obtenidos o divulgados sin previa autorización, o en ausencia de mandato legal o judicial que releve el consentimiento.

Principio de acceso y circulación restringida

La administración de los datos está sujeta a los límites que se derivan de la naturaleza de los datos, de las disposiciones de la ley y de los principios de la administración de datos personales especialmente el de temporalidad y la finalidad de la información.

Principio de temporalidad de la información

La información del titular no debe ser suministrada a terceros cuando caduque su utilidad para la finalidad del banco de datos.

Principio de interpretación integral de derechos constitucionales

Los derechos de los portadores de datos se deben interpretar en consonancia y en un plano de balance con el derecho a la información previsto en el Artículo 20 de la Constitución y con los demás derechos constitucionales aplicables como el derecho al buen nombre, el derecho a la honra, el derecho a la intimidad y el derecho a la información.

Principio de seguridad

Establece medidas que impiden el acceso a los sistemas de información por parte de personas no autorizadas para evitar el desvío de la información hacia sitios no previstos. Según la Corte Constitucional, en sentencia C-748 de 2011, se deriva entonces la responsabilidad que recae en el administrador del dato por lo que el responsable o encargado del tratamiento debe tomar las medidas acordes con el sistema de información correspondiente.

Principio de confidencialidad

Establece que todas las personas que intervengan en el tratamiento de datos personales que no tengan la naturaleza de públicos están obligadas a garantizar la reserva de la información, inclusive después de finalizada su relación con alguna de las labores que comprende el tratamiento, pudiendo solo realizar suministro o comunicación de datos personales cuando ello corresponda al desarrollo de las actividades autorizadas en la presente ley y en los términos de esta.

2.7.4 Procedimientos y requisitos para autorizar el tratamiento de datos personales:

Según el artículo 9 de la ley 1581 de 2012, sin perjuicio de las excepciones previstas en la ley, en el tratamiento de datos se requiere la autorización previa e informada del titular, la cual deberá ser obtenida por cualquier medio que pueda ser objeto de consulta posterior.

Para ello se debe seguir el procedimiento descrito en el artículo 7 del Decreto 1377 de 2013 que señala que los responsables del tratamiento de datos personales deben establecer mecanismos para obtener la autorización de los titulares o de quien se encuentre legitimado de conformidad con lo establecido en el artículo 20 del mismo decreto, que garanticen su consulta por parte del titular.

Los legitimados para obtener datos personales son los siguientes:

1. El Titular, quien deberá acreditar su identidad en forma suficiente por los distintos medios que le ponga a disposición el responsable.
2. Causahabientes del Titular, quienes deberán acreditar tal calidad.
3. Representante y/o apoderado del Titular, previa acreditación de la representación o apoderamiento.

Los responsables y encargados del tratamiento deben establecer mecanismos sencillos y ágiles que se encuentren permanentemente disponibles para los titulares con el fin de que estos puedan acceder a los datos personales que estén bajo el control de aquellos y ejercer sus derechos. Aunque la Ley 1581 de 2012 consagra explícitamente el concepto de anonimización de datos, es de considerar que la anonimización es un tipo de tratamiento de datos mediante el cual se retira cualquier identificador que pueda vincular a una persona determinada.

Dado que un dato anonimizado ya no puede vincularse a una persona natural determinada, el tratamiento de un dato anonimizado escapa en principio a la aplicación de la normatividad en materia de tratamiento de datos personales. Sin embargo, es necesario que el responsable del tratamiento de datos evalúe los riesgos de una probable reidentificación de la información o el riesgo de elaborar procedimientos de anonimización incompletos.

A continuación, se incluye el marco normativo para la protección de datos personales en Colombia. Actualmente en el marco jurídico colombiano hay obligaciones que no se encuentran reguladas, como por ejemplo el derecho al olvido, la elaboración de perfiles y la designación de delegados de protección de datos personales (GARRIGUES, 2018).

Tabla 1. Marco normativo en Colombia para la protección de datos personales

Nombre	Año	Descripción
Constitución Política de Colombia, artículo 15	**1991**	"Todas las personas tienen derecho a su intimidad personal y familiar y a su buen nombre, y el Estado debe respetarlos y hacerlos respetar. De igual modo, tienen derecho a conocer, actualizar y rectificar las informaciones que se hayan recogido sobre ellas en los bancos de datos y en archivos de entidades públicas y privadas. En la recolección, tratamiento y circulación de datos se respetarán la libertad y demás garantías consagradas en la Constitución. La correspondencia y demás formas de comunicación privada son inviolables. Sólo pueden ser interceptados o registrados mediante orden judicial, en los casos y con las formalidades que establezca la ley". (Const., 1991, pág. art 15).
Ley 1266	**2008**	Por la cual se dictan las disposiciones generales del Habeas Data y se regula el manejo de la información contenida en bases de datos personales, en especial la financiera, crediticia, comercial, de servicios y la proveniente de terceros países y se dictan otras disposiciones. (Congreso de la República de Colombia, 2008).
Ley 1273	**2009**	Por la cual se modifica el Código Penal, se crea un nuevo bien jurídico tutelado - denominado "de la protección de la información y de los datos"- y se preservan integralmente los sistemas que utilicen las tecnologías de la información y las comunicaciones, entre otras disposiciones. (Congreso de la República, 2009).
Decreto 1727	**2009**	Por el cual se determina la forma en la cual los operadores de los bancos de datos de información financiera, crediticia, comercial, de servicios y la proveniente de terceros países, deben presentar la información de los titulares de la información. (Decreto 1727 DE 2009, 2009).

Nombre	Año	Descripción
Decreto 2952	**2010**	Por el cual se reglamentan los artículos 12 y 13 de la Ley 1266 de 2008.
Ley 1581	**2012**	Por la cual se dictan disposiciones generales para la protección de datos personales. Esta ley tiene por objeto desarrollar el derecho constitucional que tienen todas las personas a conocer, actualizar y rectificar las informaciones que se hayan recogido sobre ellas en bases de datos o archivos, y los demás derechos, libertades y garantías constitucionales a que se refiere el artículo 15 de la Constitución Política; así como el derecho a la información consagrado en el artículo 20 de la misma. (Congreso de la República, 2012).
Decreto 1377	**2013**	Por el cual se reglamenta parcialmente la Ley 1581 de 2012. Con el fin de facilitar la implementación y cumplimiento de la Ley 1581 de 2012 se deben reglamentar aspectos relacionados con la autorización del Titular de información para el tratamiento de sus datos personales, las políticas de Tratamiento de los Responsables y Encargados, el ejercicio de los derechos de los Titulares de información, las transferencias de datos personales y la responsabilidad demostrada frente al Tratamiento de datos personales, este último tema referido a la rendición de cuentas.
Decreto 1074	**2015**	“Por medio del cual se expide el Decreto Único Reglamentario del Sector Comercio, Industria y Turismo.” Específicamente en el artículo 2.2.2.25.1.1. Objeto. El presente capítulo tiene como objeto reglamentar parcialmente la Ley 1581 de 2012, por la cual se dictan disposiciones generales para la protección de datos personales.

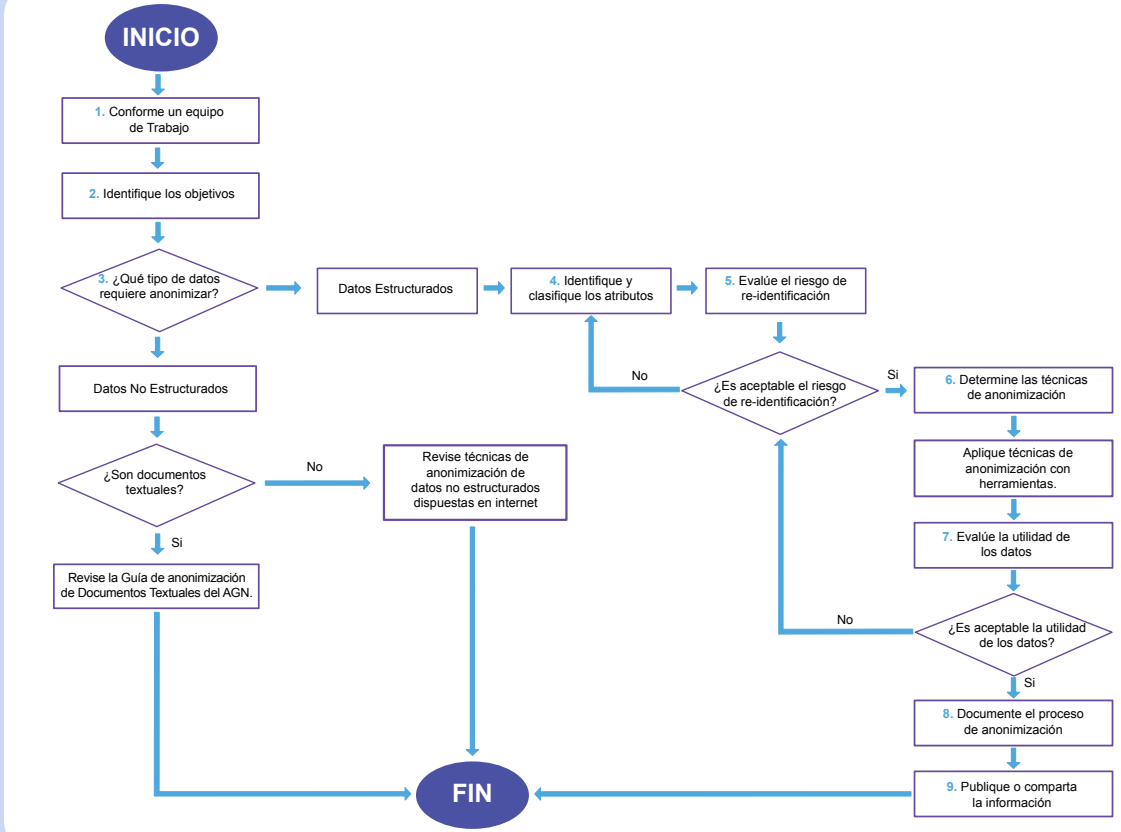
Fuente: Elaboración propia.

3. Metodología: ¿Qué pasos se deben seguir para anonimizar datos?

Volver al índice

A continuación, se propone un flujo de actividades que permiten orientar a las entidades al momento de definir su protocolo de anonimización. Se precisa que en ningún caso se trata de un modelo cerrado, sino que ofrece una estructura que puede ser tenida en cuenta en los procesos de anonimización que pretendan realizar las entidades públicas. Esta orientación metodológica tiene como alcance la anonimización de datos estructurados[1].

Ilustración 2. Diagrama metodológico para la anonimización de datos



Fuente: Elaboración propia.

[1] Los datos estructurados están organizados conforme a un modelo o esquema. Se almacenan en forma tabular y algunas veces su estructura también incluye la definición de las relaciones entre ellos. Típicamente están representados en bases de datos que hacen parte del funcionamiento de sistemas de información. (CONPES 3920).

Se refiere a aquella que está definida y sujeta a un formato concreto que facilita su procesamiento. Por ejemplo, la información organizada y estructurada en bases de datos relacionales u hojas de cálculo se considera estructurada. (MinTIC. G.INF.07 Guía Cómo construir el catálogo de Componentes de Información)

3.1 Conformación de un equipo de trabajo

Se recomienda que la entidad conforme un equipo de trabajo, el cual podría depender del área de seguridad informática, del grupo de la dirección o coordinación de tecnologías de la información o la que haga sus veces. También es importante que a este grupo se vincule el área de gestión documental o la que haga sus veces.

Lo ideal es que exista un equipo permanente dentro de la entidad que esté apoyando este proceso, sin embargo, si la entidad no tiene o puede conformar un equipo permanente, si debe conformar un equipo responsable o definir un profesional responsable para cada conjunto de datos que vaya a anonimizar. El equipo de trabajo deberá evaluar el riesgo y realizar el análisis de impacto con el fin de generar evidencias y soportes que permitan tomar decisiones ya sea en un Comité Institucional de Gestión y Desempeño o en un comité interno de anonimización, conformado por el mismo equipo de trabajo. A continuación, se sugieren los siguientes roles los cuales pueden ser desempeñados por una o varias personas:

Responsable del tratamiento de datos

Es la persona encargada de decidir la finalidad principal del uso de los datos y a los objetivos que responde el uso de la información.

Persona encargada de la protección de datos personales

La función del oficial de protección de datos o del área encargada de protección de datos en la entidad es garantizar la implementación efectiva de las políticas y procedimientos adoptados por la entidad para cumplir con la normatividad de datos personales, así como la implementación de buenas prácticas de gestión de datos personales dentro de la organización.

El oficial de privacidad tiene la labor de fomentar la cultura de datos dentro de la entidad, debe estar involucrado en el tratamiento de datos de manera oportuna y tener conocimiento sobre los sistemas de información, seguridad de los datos, normatividad en materia de protección de datos personales.

Esta persona también es la encargada de realizar evaluaciones de impacto previas sobre la privacidad de los datos, verificar la adecuada implementación de los procesos de anonimización y proveer auditorias de cumplimiento de estos.

Líder o coordinador del proceso de anonimización

Este rol es el encargado de liderar y orientar al equipo de anonimización y establecer la conexión con otras instancias de la entidad, por ejemplo, con el grupo jurídico para determinar las restricciones legales del uso de los datos, con el Comité Institucional de Gestión y Desempeño para la toma de decisiones.

Profesional o experto temático

Este rol conoce los datos que se pretenden anonimizar con el fin de asesorar en la evaluación del riesgo e impacto de desidentificación. Apoya el proceso de evaluación de la utilidad de los datos después de aplicar las técnicas de anonimización

Profesional o equipo de evaluación de riesgos

Este rol es el encargo de hacer la evaluación de riesgos, evaluar y auditar los procesos de anonimización. Este rol es el encargo de verificar que el proceso de anonimización cumple con los requisitos de un riesgo aceptable de reidentificación.

Profesional de seguridad informática y evaluación de riesgo

Es el rol encargado de realizar el análisis de riesgo de reidentificación y el impacto que pudiera existir. Así mismo es el encargado de aplicar y definir protocolos de acceso a la información por parte del equipo de trabajo que participará en el proceso de anonimización, cuando esto no exista previamente definido en la entidad. Por ejemplo, acuerdos de confidencialidad del personal involucrado en el proceso (equipo de trabajo) debidamente firmados, usos de permisos y contraseñas para el uso de la información, entre otros.

Profesional o equipo de informática

Este rol cuenta con habilidades en el manejo de herramientas de *software* para anonimización y bases de datos, es el encargado de la manipulación y procesamiento de los datos, de aplicar las técnicas de anonimización y manipular el *software* que apoya estos procesos o realizar los algoritmos que permiten aplicar las técnicas de anonimización. Se recomienda que los profesionales que asumen este rol no estén en contacto directo o administren las bases de datos o sistemas de información en la entidad, sino trabajen en ambientes replicados.

Equipo de pre-anonimización y anonimización

Este equipo se encarga de identificar las variables que se necesita anonimizar y de proponer técnicas de anonimización que deben ser validadas por el profesional o el equipo evaluador del riesgo. El equipo será el encargado de elegir y aplicar las técnicas de anonimización necesarias.

Profesional de pruebas

Este rol es el encargado de realizar las pruebas y la validación de la utilidad de los datos después de aplicar las técnicas de anonimización.

3.2 Identifique los objetivos que se quieren alcanzar con los datos anonimizados

El diseño del proceso de anonimización de datos debe estar dado por la finalidad que quiere alcanzar la entidad con el uso de la información anonimizada. La utilización de la información anonimizada puede ser en datos abiertos o con uso restringido. En este último caso, el tratamiento de datos y el proceso de anonimización puede estar acompañado de acuerdos de confidencialidad con cláusulas específicas sobre reidentificación y garantía de la privacidad de la información.

3.3 Identifique qué tipo de datos se requiere anonimizar

Se recomienda identificar si los datos son estructurados, es decir, si están organizados en forma de tablas o matrices que pueden estar en formato estructurado tal como Excel o en bases de datos relacionales que contienen los datos en tablas o, por el contrario, se trata datos no estructurados. A continuación, se incluyen las definiciones para la clasificación de los datos en estructurados, no estructurados y semiestructurados.

Características de los datos estructurados

Están organizados conforme a un modelo o esquema. Se almacenan en forma tabular y algunas veces su estructura también incluye la definición de las relaciones entre ellos. Típicamente están representados en bases de datos que hacen parte del funcionamiento de sistemas de información o están organizados en hojas de cálculo. (CONPES 3920 de 2018)

Características de los datos no estructurados

Su organización y presentación no está guiada por ningún modelo o esquema. En esta categoría se incluyen, por ejemplo, las imágenes, texto, audios, contenidos de redes sociales, videos, documentos en Word o pdf. (CONPES 3920 de 2018).

Características de los datos semiestructurados

Su organización y presentación tiene una estructura básica (etiquetas o marcadores), pero no tiene establecida una definición de relaciones en su contenido. En esta categoría se incluyen contenidos de e-mails, tweets, archivos XML (CONPES 3920 de 2018)

3.4 Identifique y clasifique los atributos

En este paso se clasifican los atributos del conjunto de datos como identificadores directos, indirectos o no identificadores. Esto determinará la forma en que a los atributos se les dará el tratamiento y procesamiento posterior.

Existen tres condiciones para que una variable se considere un identificador (HHS, 2012):

I Replicabilidad

Los valores del campo deben tener cierta consistencia a lo largo del tiempo para que se pueda hacer un análisis de consistencia en relación con el sujeto. Si un valor de un campo no es consistente y replicable en el tiempo será muy difícil reidentificar al sujeto. Un ejemplo de esto son los resultados de glucosa en la sangre de un paciente, los cuales no se mantienen a lo largo del tiempo, pero en cambio el factor Rhesus (RH) es invariable a lo largo de la vida de una persona.

II Distinguibilidad

Los valores del campo deben tener suficiente variación para distinguir al sujeto de otros valores, dentro de un conjunto de datos. Ejemplo de valores distinguibles en un conjunto de datos son el número de identificación, el teléfono celular, la dirección, entre otros.

III Conocible

Para poder reidentificar a un sujeto, es necesario que se conozca previamente algunos de los identificadores y variables asociadas a este. Cuando se habla de una variable conocida, significa que el atacante[2] tiene una identidad adjunta a esa información, como por ejemplo el código postal, la fecha de nacimiento y el nombre del sujeto. Es decir, que el atacante tiene conocimiento previo de esta información, especialmente si éste es cercano al sujeto que se quiere identificar, como por ejemplo un vecino o un compañero de trabajo.

[2]. En esta guía se denomina atacante a la persona que intenta hacer el proceso de reidentificación del sujeto. Otros términos relacionados son adversario o intruso

Por el contrario, un no conocido tiene una probabilidad de conocimiento menor sobre las variables del sujeto. Sin embargo, pueden ser inferidas a partir de otras variables conocidas, así mismo puede tener conocimiento de otras variables si parte de la información del sujeto ha sido publicada en algún medio físico (periódicos, revistas) o electrónico (redes sociales, portales). También se puede saber si la información es semipública o si la información se puede comprar a través de bases de datos comerciales.

Una vez se tiene definido si la variable es un identificador, entonces se procede a clasificarlo como identificador directo o identificador indirecto. Se debe tener en cuenta que existen identificadores directos e indirectos de tipo clasificado, que consisten en características de datos que pueden afectar la intimidad del titular y su uso indebido puede generar violaciones a derechos, tales como el origen étnico, la orientación política, las convicciones religiosas o filosóficas, la pertenencia a sindicatos u organizaciones sociales, y los relativos a la salud, la vida sexual y los datos biométricos.

Si los atributos son identificadores, estos pueden ser directos o indirectos:

* **Identificadores directos**

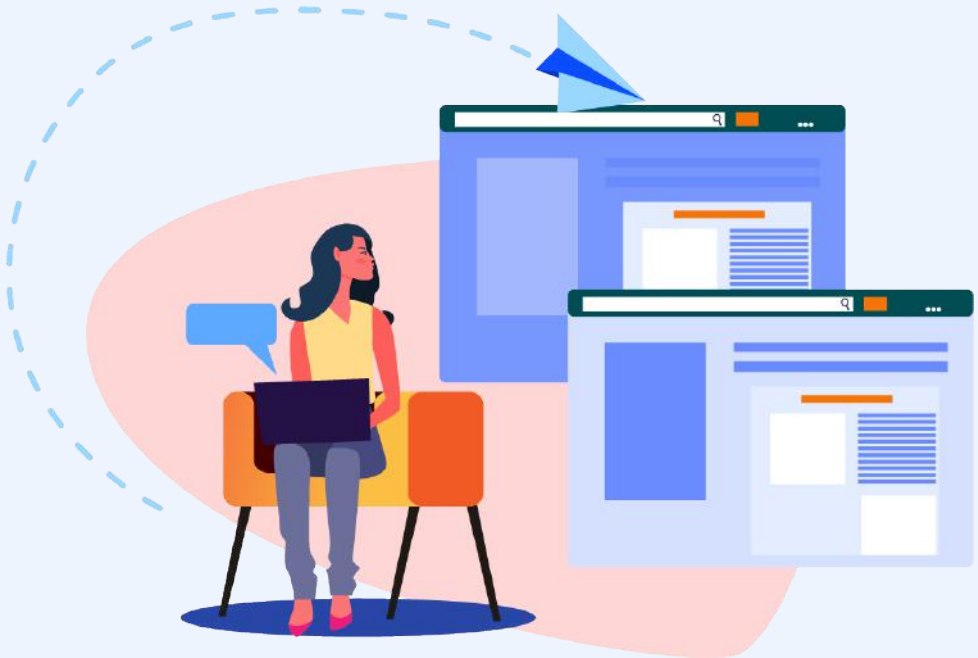
Son todas aquellas características o atributos distintivos que por sí mismos permiten la identificación de una persona natural o jurídica de manera inequívoca dentro de un conjunto de datos. Los identificadores directos tienen dos atributos principales:

1. Uno o más identificadores se pueden usar para identificar un individuo ya sea por si mismos o en una combinación con otra información.

2. Usualmente no son útiles para el objetivo del análisis.

Algunos ejemplos son:

- Identificador único (número de cedula, número de documento de residente, número de pasaporte, número de documento extranjería y el número de licencia de conducción).
- Dirección detallada.
- Número de teléfono (celular, casa, oficina, fax, etc.).
- Número de documento médico, número del beneficiario de la asistencia social.
- Número de cuenta bancaria, número de tarjeta de crédito.
- Número de matrícula, y número de documento y número de serie de diferentes equipos.
- Fotos (fotos fijas, video, videos de CCTV, etc.).
- Datos biométricos (huellas dactilares, voz, iris, etc.).
- Dirección de correo electrónico, dirección IP, dirección de Mac, URL de la página de inicio, etc.
- Código de identificación (identificación, número de empleado, número de cliente, etc.).
- Número de la tarjeta militar.
- Número de tarjeta profesional.
- Otros diferenciales como los enmarcados en la “Regla de Privacidad HIPAA” emitida por el Departamento de Salud y Servicios.



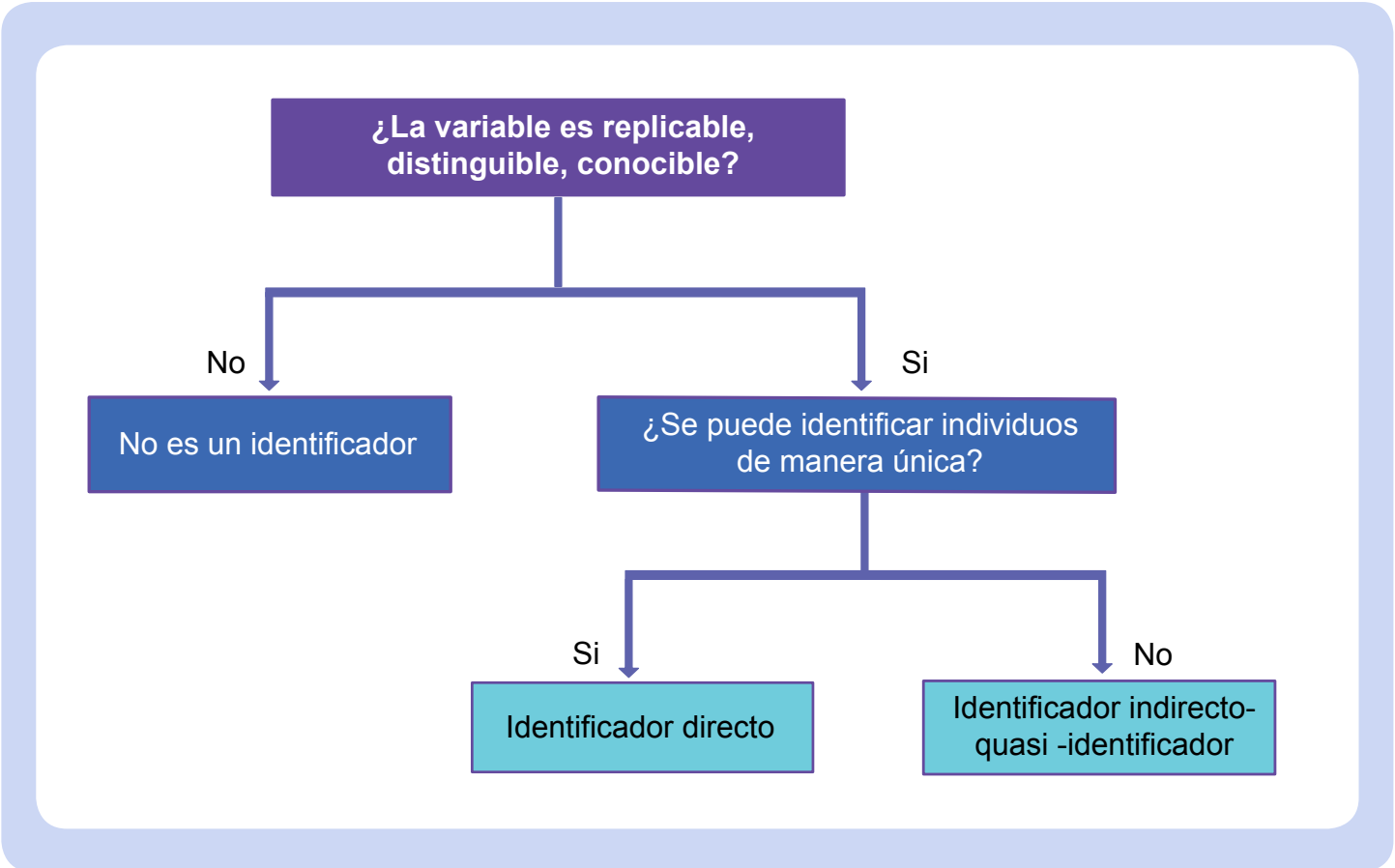
* **Identificadores indirectos o cuasi - identificadores**

Son aquellas características o atributos que por sí solas no permiten la identificación de una persona natural o jurídica, pero que relacionados o en combinación con otros identificadores indirectos podrían permitir la identificación dentro de un conjunto de datos. El que una persona natural sea identificable o no, con base en la combinación de identificadores indirectos, también puede depender del dominio específico. Por ejemplo, la combinación de los atributos **"mujer"**, **"45"** y **"abogada"** puede bastar para identificar a una persona dentro de una compañía particular; sin embargo, a menudo será insuficiente para identificar a esa persona natural fuera de la compañía.

Son ejemplos de identificadores indirectos la fecha de nacimiento, código postal, datos de la fecha de aniversario, años de escolaridad.

En la **Ilustración 3**, se resumen el proceso de clasificación de los atributos.

Ilustración 3. **Ruta para la identificación y clasificación de los atributos**



Fuente: Elaboración propia, adaptado de Sharing Clinical Trial Data. Concept and methods for de-identifying Clinical Trial Data. (Emam & Malin, 2014)

3.5 Evalúe el riesgo de reidentificación

El riesgo de anonimización es la probabilidad de reidentificar a un individuo u organización dentro de un conjunto de datos. Es importante tener en cuenta que ninguna técnica de anonimización podrá garantizar efectividad absoluta, ya que existirá siempre un índice de probabilidad de reidentificación que se debe intentar reducir mediante la correspondiente gestión de riesgos.

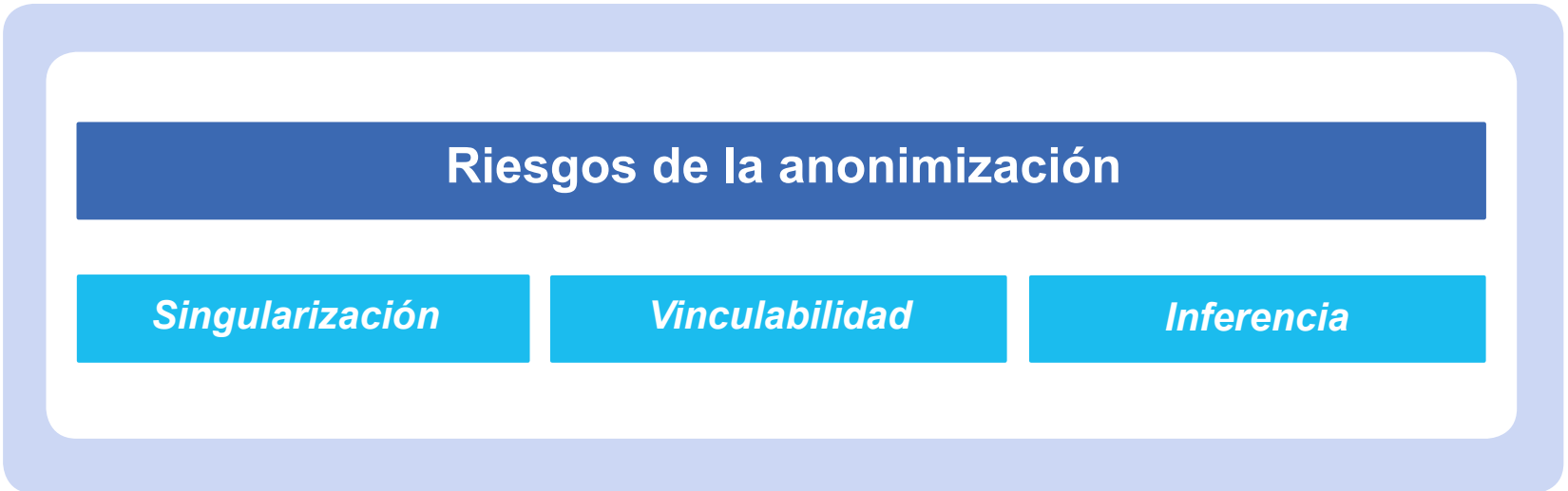
El riesgo de reidentificación está implícito y aumenta a medida que transcurre el tiempo desde la anonimización de los datos, como consecuencia de la evolución e incremento de los identificadores indirectos a lo largo del tiempo como, por ejemplo, la información que el propio interesado haya aportado sobre sí mismo en redes sociales, portales web, blogs, etc. Para la medición del riesgo de reidentificación es necesario establecer los umbrales para determinar si el riesgo es alto, medio y bajo. Generalmente se calcula usando modelos de probabilidad en el que se establecen los riesgos y probabilidades de ocurrencia que parten de los factores de daño que se causarían a los titulares de los datos en caso de reidentificación, así como las consecuencias económicas para la entidad que los publica.

Aunque no es objeto de esta guía, se recomienda que la entidad antes de realizar la evaluación de los riesgos de anonimización y reidentificación, elabore una Evaluación de impacto en la privacidad con el fin de determinar el riesgo que genera el tratamiento de información personal para los titulares de los datos. La evaluación de impacto de datos personales permite identificar de manera previa aquellos datos que son estrictamente necesarios para el objeto que se quiere alcanzar y permite planificar la gestión de los riesgos de la información y los recursos que se utilizarán para este fin, de acuerdo con la normatividad vigente.

La evaluación de riesgos de reidentificación puede ser comprendida como una actividad dentro de la política de evaluación de impacto de privacidad implementada por la entidad.

A continuación, se enumeran los principales aspectos que los responsables de la anonimización y del tratamiento del riesgo deben tener en cuenta a la hora de considerar aplicar alguna técnica. De igual forma hay que valorar, en particular, la garantía que se obtiene al aplicar una determinada técnica, teniendo en cuenta el estado actual de la misma y los tres riesgos principales de la anonimización:

Ilustración 4. Ruta para la identificación y clasificación de los atributos



Fuente: Elaboración propia

A continuación, se explica cada una de ellas.

a Singularización

Ocurre cuando es posible distinguir, extraer o particularizar los datos relacionados que identifican a una persona dentro de un conjunto de datos o de algunos datos (o todos los datos). Esto puede deberse a que la información relacionada con un individuo tiene un valor único y se podría identificar, distinguir o particularizar dentro de un conjunto de personas.

Ejemplo: en un conjunto de datos que registra la altura de los individuos, donde solo una persona mide 2,5 metros de alto, ese individuo se destaca. También puede ocurrir si se conectan datos diferentes relacionados con las mismas personas en el conjunto de datos y una persona tiene una combinación única de valores. Por ejemplo, podría haber solo un individuo en un conjunto de datos que mide 1,40 cm de alto y nació en 1990, aunque hay muchos otros que comparten la altura o el año de nacimiento.

Tabla 2. Ejemplo registros de datos de individuos

Nombres	Apellidos	Estatura	Edad	Lugar de nacimiento
Juan	**Pérez**	2,50 cm	25	Bogotá
Pedro	**Páez**	1,65 cm	26	Bogotá
Lucas	**Pinto**	1,40 cm	24	Girardot
Mateo	**Gómez**	1,65 cm	25	Bogotá

Fuente: Elaboración propia

b Vinculabilidad

Cualquier enlace de datos o identificadores en un conjunto de datos hará que sea más probable que un individuo sea identificable. Este riesgo ocurre cuando se pueden vincular como mínimo dos datos de una única persona o de un grupo de interesados, ya sea en la misma base de datos o en dos bases de datos distintas. No basta con aislar los datos de un individuo de una base de datos si al compararla con otra permite identificarlo respecto a un grupo.

Ejemplo: si se toma individualmente el primer y segundo nombre “Sebastián” y “Yatra” puede que no sea capaz de distinguir a uno de los clientes de una empresa grande de todos los demás clientes, pero si las dos piezas de información están vinculadas, es mucho más probable que “Sebastián Yatra” se referirá a un individuo único e identificable. Cuantos más identificadores estén vinculados entre sí en un conjunto de datos, mayor será la probabilidad de que la persona con quien se relaciona sea identificada o identificable.

Si se puede determinar (p. ej., mediante un análisis de correlación) que dos datos están asignados al mismo grupo de personas, pero no puede singularizar a las personas en este grupo, entonces la técnica es resistente a la singularización, pero no a la vinculabilidad.

c Inferencia

Consiste en inferir o deducir un vínculo entre dos elementos de información en un conjunto de datos, aunque la información no esté explícitamente vinculada.

Ejemplo: si un conjunto de datos contiene estadísticas sobre la antigüedad (el tiempo en que el trabajador ha prestado servicios para una empresa) y el pago de los empleados. Si bien estos datos no apuntarían directamente a los salarios de los individuos en el conjunto de datos, se podría hacer una inferencia entre los dos datos, lo que permitirá identificar a algunos individuos. Existe un riesgo de reidentificación que debe ser considerado por las entidades puedan salvaguardarse adecuadamente.

Cada organización deberá establecer políticas y procedimientos que definan las condiciones de acceso a los datos y definir estrategias en aras de prevenir estos tres riesgos e impedir la reidentificación e imposibilitar el mal uso por parte del responsable del tratamiento o de cualquier tercero.

Es importante señalar sobre las técnicas de desidentificación y anonimización que no hay una técnica cien por ciento (100%) segura.

3.5.1 Metodolgía de medición del riesgo

En esta guía se propone el siguiente método basado en la Guía de Anonimización del gobierno de Canadá. Este contempla una medición del riesgo de carácter cuantitativo que incorpora a su vez insumos cualitativos. Es importante mencionar que el cálculo del riesgo se calcula antes de aplicar las técnicas de anonimización tal y como se menciona en la **ilustración 2.**

3.5.1.1 Evaluar el nivel de invasión de la privacidad de un conjunto de datos:

Para ello se debe suponer que la información en el conjunto de datos es identificable. En esta evaluación se debe tener en cuenta:

- La sensibilidad de la información.
- El nivel de detalle de la información.
- El número de individuos.
- El daño potencial que se puede ejercer sobre una persona por el uso inadecuado de su información.
- Si la información no fue solicitada o dada libremente por los individuos, con poca o ninguna privacidad.
- Si las personas dieron su autorización para que su información fuese difundida y/o se les notificó antes de revelarla.

El resultado de la evaluación de los aspectos anteriormente descritos es de carácter cualitativo, para lo cual se podrá realizar una tabla donde cada uno de los aspectos se evalúe por cada uno de los miembros del equipo de anonimización, en una escala de 1 a 5 donde 1 es nivel de invasión bajo y 5 nivel de invasión muy alto, y al final se promedia y se determina el nivel de invasión de la privacidad de manera cualitativa. En conclusión es necesario implementar una metodología cuantitativa que permita clasificar los riesgos y medir los resultados con valores numéricos.

3.5.1.2 Establecer un umbral de riesgo

Representa el valor máximo de riesgo relacionado con los datos y los problemas de reidentificación. Esta medida se establece en rangos de 0 a 1 y refleja el nivel de riesgo de reidentificación máximo aceptado. El umbral del valor de riesgo tiene relación directa con el daño potencial que ocasione la reidentificación del sujeto titular de los datos. Se debe tener en cuenta, que nunca se puede garantizar una probabilidad de reidentificación de 0. El umbral de riesgo va a definir las transformaciones que se deben hacer en el sistema de datos para cumplir con el valor máximo de riesgo aceptado.

3.5.1.3 Medición de la cantidad de riesgo de reidentificación

Para continuar con la evaluación es necesario calcular **i)** la probabilidad de reidentificación para un solo registro o individuo dentro del conjunto de datos, **ii)** aplicar el método de medición de riesgo apropiado.

Probabilidad de reidentificación: cada fila de un conjunto de datos contiene información de un individuo. En ese sentido, cada fila tiene una probabilidad de reidentificación que depende de cuantas otras filas del conjunto de datos tienen los mismos valores o cuasidentificadores, la probabilidad, por tanto, depende del número de personas que tienen los mismos valores para cada uno de los datos. El cálculo para identificar la probabilidad, se hace de la siguiente forma: de revelarla.

$$\text{Probabilidad de reidentificación} = \frac{1}{\text{(Número de personas con mismos valores)}}$$

Para ejemplificar el proceso supóngase que en un colegio hicieron una lista de veinte estudiantes con las mejores notas, la lista está organizada por orden alfabético para que los profesores puedan tener idea del desempeño de sus estudiantes. Sin embargo, para no generar problemas entre los estudiantes, la tabla publicada no contiene nombres solo contiene el sexo del estudiante, su nota y curso.

Tabla 3. Ejemplo probabilidad de reidentificación

Sexo	Nota	Curso
Mujer	5	11
Hombre	4.9	11
Mujer	5	11
Mujer	4.8	11
Hombre	5	10
Mujer	4.8	11
Hombre	4.9	10
Hombre	4.8	10
Mujer	5	11
Hombre	4.8	10
Mujer	5	11
Mujer	4.9	11
Hombre	5	11
Mujer	4.8	10
Hombre	5	10
Mujer	5	11
Hombre	5	10
Hombre	4.8	10
Mujer	5	11
Hombre	4.8	11

i

Ahora se evalúa la probabilidad de reidentificación para las mujeres de grado once con nota de cinco y para los hombres de décimo grado con nota de cinco.

$$\text{Reidentificación mujer de 11 con nota 5} = \frac{1}{\text{\# de mujeres de 11 con nota 5}} = \frac{1}{6} = 0{,}16666$$

$$\text{Reidentificación hombre de 10 con nota 5} = \frac{1}{\text{\# de hombres de 10 con nota 5}} = \frac{1}{3} = 0{,}16666$$

La probabilidad de reidentificación de una mujer de grado once con nota de cinco es de 0,16666, mientras que para un hombre de décimo grado con nota de 5 la probabilidad es de 0,33333. Por tal razón los hombres tienen un mayor riesgo de reidentificación, debido a que hay un menor número de participantes que cumple con tales condiciones.

ii

Aplicar el método de medición de riesgo adecuado.

Una vez se ha medido la probabilidad de reidentificación para cada fila de datos, es necesario identificar cuál es el mejor modelo para garantizar la no reidentificación de los datos. Para poder establecer las medidas de riesgo es necesario identificar el contexto de la base de datos y, por tanto, se debe establecer si la información es pública, no pública o semi pública. Dependiendo de estos tres contextos se podrá identificar el tipo de agentes que pueden acceder a la información y por tanto el tipo de usos que le pueden dar.

- **Publicaciones de datos públicos:**

En el caso de datos públicos, es necesario pensar en una probabilidad alta de ataque el cual apunta a las filas más vulnerables donde hay clases de equivalencia más pequeñas y por consiguiente mayor probabilidad de reidentificación. En este caso, se debe asignar la probabilidad máxima de reidentificación en todas las filas; se generaliza, para toda la tabla, la probabilidad de riesgo del individuo con mayor posibilidad de reidentificación.

- **Publicaciones de datos no públicos:**

En este caso, el acceso a los datos es de carácter limitado y en ese sentido se debe aplicar la probabilidad

ii

de reidentificación media en todas las filas (es decir 0,5), asumiendo que ninguna fila es más vulnerable que otra frente a un ataque de reidentificación. Para proteger filas únicas o clases de equivalencia con más probabilidad de reidentificación, es necesario identificar un promedio estricto entre todos las filas, donde ninguna fila tenga una probabilidad de reindentificación mayor a un valor específico.

- **Publicaciones de datos semipúblicos:**

Al igual que en el primer paso, es necesario pensar en una probabilidad alta de ataque, por consiguiente, se debe asignar la probabilidad máxima de reidentificación en todas las filas (es decir 1), que depende del individuo con mayor probabilidad de reidentificación.

3.5.1.4 Medir el riesgo según el contexto

Además del riesgo de reidentificación en los datos, el riesgo de un conjunto de datos está dado por los tipos de ataques a los conjuntos de datos según el modelo de publicación o lanzamiento que se emplee. El análisis más detallado del riesgo de identificación en función de posibles ataques se denomina riesgo de contexto.

El riesgo de contexto es la probabilidad de que se realicen uno o varios ataques de reidentificación en contra del conjunto de datos una vez que se ha publicado. Los atacantes y sus tipos de ataques dependen del modelo de lanzamiento que se ha utilizado. El riesgo de contexto junto al riesgo de datos (explicado anteriormente) permite calcular el riesgo general de reidentificación en la publicación de un conjunto de datos.

A continuación, se menciona la probabilidad de reidentificación de contexto de acuerdo con la publicación de datos públicos, semipúblicos y no públicos.

a Publicación de datos públicos

Dado que el conjunto de datos está disponible para ser descargado por cualquier persona y sin ninguna condición, se debe asumir que la probabilidad de ataques de reidentificación contra el conjunto de datos, es de 1.

b **Publicación de datos no públicos**

En este caso, las técnicas para estimar la probabilidad son más complejas y requieren de conocimientos especializados para obtenerlos. Para este modelo se recomienda medir el riesgo de contexto como si se tratara de una publicación de datos públicos, utilizando el método anterior.

Para la estimación de la probabilidad de ataque de publicación de datos no público, se deben determinar las probabilidades de tres ataques de reidentificación clasificadas en tres tipos:

- Ataque interno deliberado.
- Reconocimiento involuntario de un individuo, en el conjunto de datos.
- Violación de datos.

***** **Ataque interno deliberado**

La probabilidad de que se genere un ataque interno deliberado depende de dos factores. Por un lado, los mecanismos y acuerdos de seguridad y privacidad relacionados con el uso o manejo de los datos no públicos. Por otro lado, los motivos y capacidades del destinatario para realizar un ataque de reidentificación. La unión de estos dos factores da como resultado nivel de riesgo bajo, medio o alto de reidentificación.

En el caso de los controles de seguridad y privacidad, dependiendo de la complejidad del sistema y la capacidad de monitoreo (bajo, medio o alto), será más difícil o fácil para el destinatario realizar ataques de reidentificación. Por su parte, los motivos y capacidades relacionados con riesgos de reidentificación se relacionan con: problemas entre el destinatario de la información y la organización; la existencia de incentivos, económico o de otro tipo, que impulsen a destinatario de reidentificar a uno o más individuos de la base de datos; ventajas del destinatario en términos de conocimientos y recursos técnicos y tecnológicos necesarios para llevar a cabo ataques de reidentificación; y, acceso a otras bases de datos que permitan relacionar la información de una o más personas (Information and Privacy Commissioner of Ontario , 2016).

Con base en lo anterior, cada organización debe calcular el nivel de riesgo de ataque deliberado, conjugando los niveles de seguridad y privacidad con los motivos y capacidades de reidentificación. A continuación, se presentan los posibles resultados.

Tabla 4. Probabilidad de reidentificación ataque interno deliberado

Controles de privacidad y seguridad	Motivos y capacidades	Probabilidad de ataque interno deliberado
Alto	Bajo	0,05
	Medio	0,1
	Alto	0,2
Medio	Bajo	0,2
	Medio	0,3
	Alto	0,4
Bajo	Bajo	0,4
	Medio	0,5
	Alto	0,6

Tabla de elaboración propia con información de (Information and Privacy Commissioner of Ontario , 2016)

*** Reconocimiento involuntario de un individuo, en el conjunto de datos:**

En cálculo del riesgo por reconocimiento involuntario, se asume la probabilidad de que el destinatario de la información no pública, reidentificación a un individuo de los sistemas de información de manera fortuita, ya sea porque es un amigo, un familiar, conocido, etc. Para calcular el riesgo de reconocimiento involuntario se hace uso de la siguiente fórmula:

$$1 - (1-p)^m \quad (1)$$

En la ecuación (1), **p** representa el porcentaje de personas que cumplen con la misma característica o condición, por su parte, **m** representa el número de personas, dentro del porcentaje **p**, que pueden ser reconocidas por el destinatario de la información. La variable **m**, comúnmente se asocia al número Dunbar's (150), que representa un límite teórico de individuos que son conocidas o sostienen una relación con una persona (Information and Privacy Commissioner of Ontario , 2016).

* Violación de datos

El riesgo de violación de datos ocurre cuando el destinatario de la información no pública incumple los acuerdos de privacidad y seguridad de la información. Bajo este escenario se debe asumir la posibilidad de que se difunda la información por fuera de los límites establecidos y, por tanto, que se pueda dar un ataque de reidentificación dirigido por agentes externos. En consecuencia, para calcular el valor de riesgos de violación de datos, se debe hacer uso de la información disponible sobre violación de datos en el sector asociado al destinatario de la información no pública (Information and Privacy Commissioner of Ontario, 2016).

C Publicación de datos semipúblicos

En este caso, las técnicas para estimar la probabilidad pueden considerarse iguales a las empleadas en contextos de publicación de datos no públicos incluyendo un ajuste en relación con el ataque interno deliberado.

* Cálculo total del riesgo

El cálculo total de riesgo se identifica una vez el riesgo de los datos y el riesgo del contexto han sido medidos. El riesgo total es igual al riesgo de los datos multiplicado por el riesgo del contexto. De esta forma, el riesgo total se interpreta de la siguiente forma: es la probabilidad de que una o más filas sean reidentificadas si se lanzó un ataque de reidentificación.

Riesgo total: Probabilidad de riesgo de los datos* Probabilidad de riesgo del contexto

Es entonces la unión entre la probabilidad de reidentificación y el contexto de la información la que nos permite determinar el riesgo de reidentificación en su conjunto, ya que tendremos la información suficiente para saber que agentes pueden acceder a los datos y que tan fácil es asociar directamente la información con los sujetos caracterizados.

* Transformación de la presentación de los datos

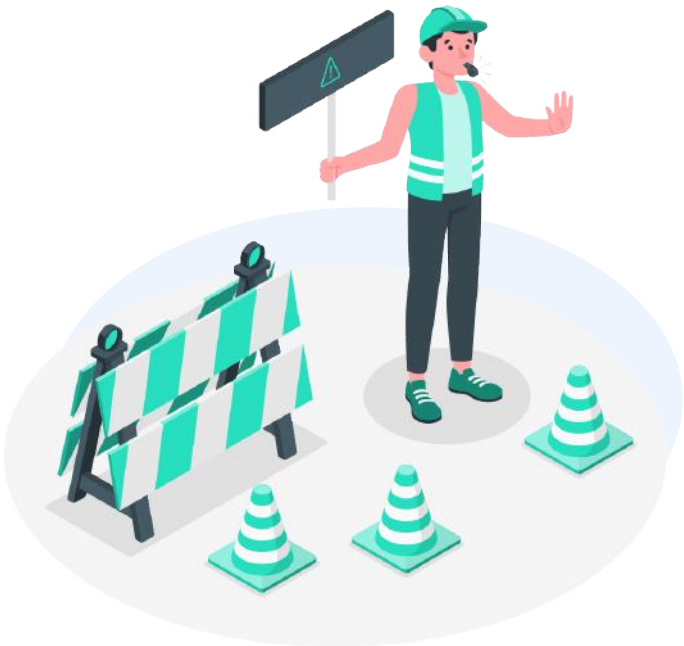
Una vez se ha estimado el cálculo total de riesgo, el último paso que se debe llevar acabo es depurar y arreglar la información para que el riesgo total este dentro de los rangos establecidos en el umbral de riesgo, que se determina en las primeras etapas del ejercicio. La adecuación de los datos y su representación solo se debe llevar acabo si el riesgo total supera el umbral de riesgo, para minimizar el riesgo total se pueden tomar diferentes acciones, como, por ejemplo: eliminar datos irrelevantes o agrupar datos por rangos de valor.

3.5.2 Gestión de información para mitigar el riesgo

- Se debe tener claridad acerca de la finalidad de la publicación o del intercambio de información.
- La entidad debe tener claramente señalados los niveles de acceso, las responsabilidades y la circulación interna y externa de la información sin anonimizar y la anonimizada.
- La entidad debe garantizar la formación y la información con la que cuenta el equipo de trabajo encargado de los procedimientos.
- Siempre se debe calcular el riesgo teniendo en cuenta que puede existir interés de identificar a un individuo dentro de un conjunto de datos anonimizados ya publicado.
- Se debe calcular el riesgo independientemente del tipo de datos anonimizados que se publica y de que se identifique un posible interés en reidentificar.
- A continuación, se menciona la probabilidad de reidentificación de contexto de acuerdo con la publicación de datos: públicos, semipúblicos y no públicos.

Los responsables del tratamiento deben tener en cuenta que un conjunto de datos anonimizado puede poseer riesgos residuales para los interesados. Los riesgos residuales son aquellos que subsisten, después de haber implementado controles y aplicado técnicas de anonimización.

Una vez identificado el nivel de riesgo de reidentificación en los resultados finales del proceso de anonimización, se debe evaluar periódicamente los impactos de la reidentificación.



Se deben tener en cuenta los siguientes aspectos dentro de la evaluación de los riesgos:

- Riesgos originados en la aplicación inadecuada de los procedimientos y las técnicas.
- El riesgo se puede incrementar al agregar más datos a lo largo del tiempo o como resultado del desarrollo de programas o proyectos específicos (por ejemplo: la atención a víctimas, la ampliación del cubrimiento en educación, etc.).
- El riesgo se incrementa en cuanto a los identificadores indirectos, por parte del titular de la información por medio del uso de redes sociales y otros medios de comunicación.
- Los riesgos son susceptibles a los cambios a los procesos o técnicas de anonimización empleadas por lo que se recomienda hacer una evaluación periódica en todo el ciclo de vida de la información. Esta revisión periódica permitirá que los evaluar el estado real de los riesgos.

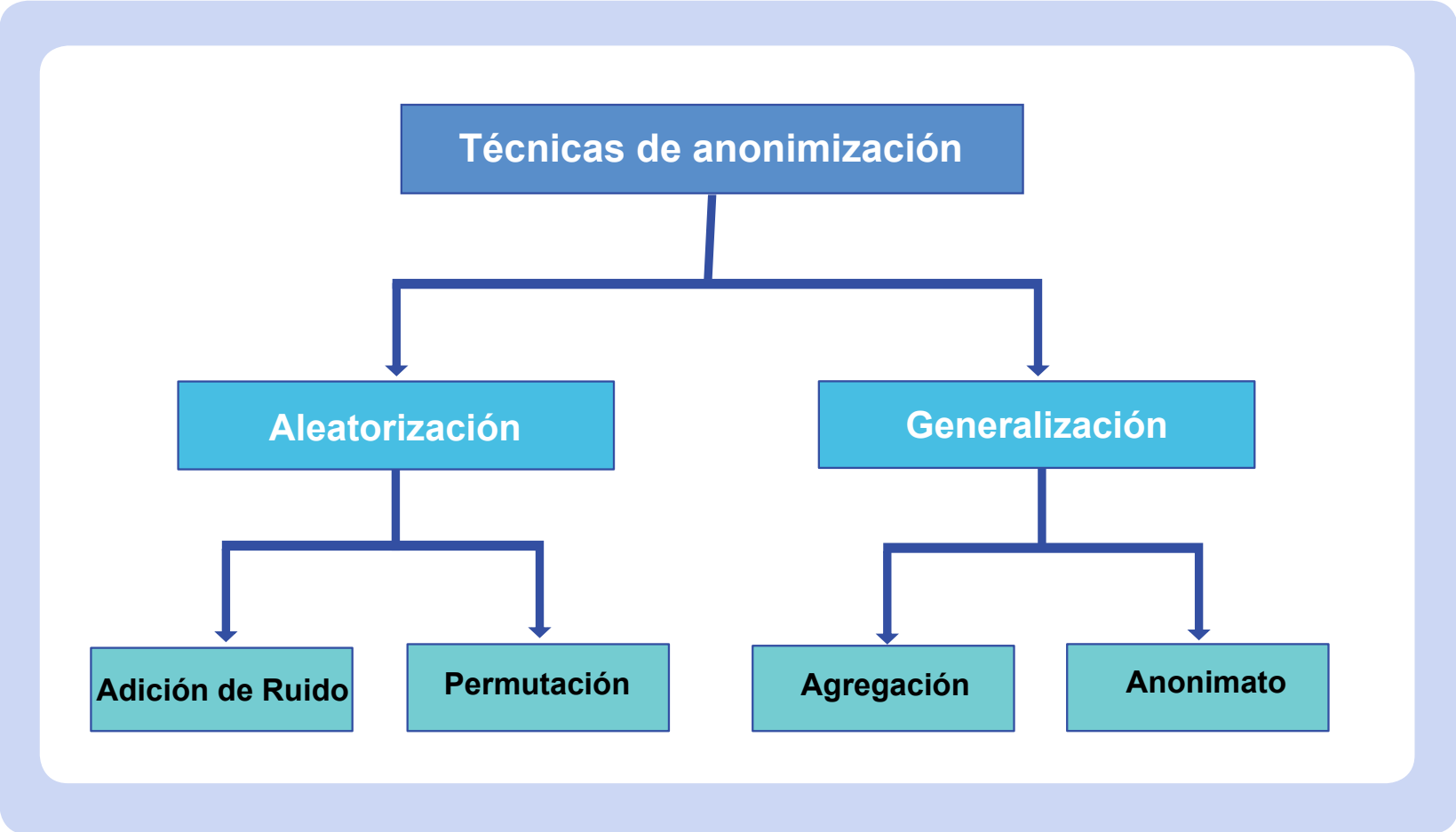
3.6 Determine las técnicas de anonimización

Las técnicas de anonimización dependen del tipo de identificadores encontrados en el paso 4 (de la gráfica 2) y del uso que se le va a dar a los datos. Se recomienda que la determinación de las técnicas de anonimización sea realizada por el equipo de trabajo, así como la realización de pruebas que permitan verificar su efectividad.

Existen limitaciones inherentes a las técnicas de anonimización. Los responsables del tratamiento deben ponderar seriamente estas limitaciones, junto con el propósito por el cual se quiere publicar los datos, antes de seleccionar una técnica u otra. Asimismo, deben atender a los fines previstos para la anonimización, como proteger la privacidad de las personas cuando se publica un conjunto de datos o permitir que se consulte algún tipo de información contenida en dicho conjunto de datos.

En términos generales, existen tres enfoques diferentes de técnicas de anonimización: el primero se basa en la aleatorización, el segundo en la generalización y el tercero en la seudonimización.

Ilustración 5. Técnicas de anonimización



Fuente: Elaboración propia tomada de Luk Arbuckle, Privacy Analytics

3.6.1 Aleatorización

Consiste en una agrupación de técnicas que alteran o modifican la autenticidad y la veracidad de los datos para eliminar el vínculo entre ellos y una persona en concreto, es decir, que los datos por sí solos sean lo suficientemente inciertos para que no sea posible volverlos a vincular entre ellos y/o con un individuo específico.

En este sentido, no se pierde la singularidad de los datos, ya que estos pueden obtenerse a partir de un único interesado, pero sí pueden protegerse contra riesgos de inferencia o deducción.

Esta técnica puede combinarse además con técnicas de generalización para obtener niveles de privacidad más altos, con el fin de evitar abstracciones realizadas a partir de un conjunto de datos que permitan plantear formulaciones o afirmaciones, posiblemente también sea necesario aplicar otras técnicas que garanticen que un registro no sea útil para identificar a una persona, pero si lo sea una vez sea anonimizado.

Dentro de la aleatorización existen varias modalidades:

a Adición de ruido

Las propiedades del conjunto de datos son transformadas haciéndolos menos precisos, pero si conservando su distribución general, de este modo un observador supondrá que los valores son idénticos, lo que será cierto hasta un punto (Comisión Europea, 2014). Aplicada de manera correcta no permite que se pueda identificar a un individuo, tampoco reparar los datos o detectar cómo se han modificado. Esta técnica es útil cuando los atributos pueden causar un importante efecto adverso en las personas.

Como ejemplo, para describir la altura de una persona que mide 1.60 m, no se suministra el dato exacto, se presenta con una diferencia de la altura de ± 10 cm, es decir 1.50 ó 1.70. Así para un tercero se dificultará o limitará la posibilidad de identificar a una persona, restaurar los datos y/o averiguar cómo se han modificado.

Para maximizar su eficacia considerando la cantidad, tipo de información y el impacto que produciría la revelación de los atributos a proteger sobre la privacidad de las personas, se recomienda combinar esta modalidad con otras técnicas de anonimización, como los cuasi identificadores o la eliminación de atributos obvios.

Algunos errores al aplicar adición de ruido

- **Agregar ruido inconsistente:** si el ruido no presenta coherencia semánticamente, es decir, está fuera de escala y entre los atributos de un conjunto de datos no se presenta una lógica congruente Por lo que facilita eliminar el ruido y facilita la recuperación de los atributos que faltan, si existen pocos elementos en el conjunto de datos, posibilita la vinculación de las entradas de datos con ruido frente a una fuente externa.
- **Asumir que es suficientes con la adición de ruido:** jamás debería pensarse que esta es una solución completa de anonimización, esto se daría si es mayor el ruido que información contenida en el conjunto de datos.

Inconvenientes de la adición de ruido:

En determinados casos se producen defectos en vez de errores, tal es el caso, por ejemplo, de la realización de una reidentificación que se lleva a cabo en una base de datos de clientes del proveedor de contenidos de videos.

Los técnicos analizaron las propiedades que tiene la base de datos y la anonimizaron. La empresa la hizo pública, teniendo en cuenta la normativa de protección de datos. Para ello lo que procedió a hacer fue eliminar todo tipo de información que pudiera identificar al cliente, excepto las valoraciones y las fechas. Se añadió ruido a las valoraciones mejorándolas o empeorándolas ligeramente.

A pesar de ello, se descubrió que se podía identificar de manera unívoca el 99 % de los registros de usuarios en el conjunto de datos usando 8 valoraciones y fechas con errores de 14 días a modo de criterio de selección. Aun rebajando los criterios de selección a 2 valoraciones y un error de 3 días, se podía identificar al 68 % de los usuarios. (Comisión Europea, 2014).

b Permutación

Esta técnica consiste en mezclar los valores de los atributos en una tabla para que puedan vincularse de manera artificial a distintos interesados (AGESIC, 2017). Esta técnica es pertinente en el caso en que quiera conservarse la distribución exacta de los datos. Se caracteriza por el intercambio de los valores incluidos en un conjunto de datos a partir del traslado de datos de un registro a otro.

El principal riesgo de esta técnica es aplicarla sobre atributos que son altamente correlacionados. Por ejemplo, intercambiar valores entre personas, de variables como año de nacimiento, años de experiencia e ingresos recibidos. En este caso, puede ser fácil reidentificar a un sujeto porque las tres variables tienen estrecha relación entre sí.

De igual forma que la aleatorización, la permutación por si sola no permite garantizar un alto nivel de anonimización y por tanto es necesario que se aplique con otras técnicas.

La permutación aplicada por si sola permite identificar los registros de la persona, puede facilitar una vinculación incorrecta entre registros, asociando a otro sujeto un valor real en un atributo que ha sido previamente intercambiado. Así mismo, es posible hacer inferencia del sujeto, especialmente si los atributos tienen una correlación estrecha.

Algunos errores al aplicar la permutación:

- **Selección del atributo equivocado:** es posible que se aplique la técnica de permutación en atributos que no impliquen un riesgo para la reidentificación, por tanto, su aplicación, no conllevaría a ningún beneficio en la protección de datos personales
- **Permutar los atributos de manera aleatoria:** aplicar la permutación sobre los atributos fuertemente correlacionados, contribuye a aumentar la probabilidad de reidentificación.

3.6.2 Generalización

El proceso o enfoque de generalización tiene como propósito modificar datos a través de escalas u órdenes, para generar esquemas de datos de acuerdo con características comunes. Una situación que puede ejemplificar el proceso de generalización es transformar el dato de una ciudad por el departamento o región de la cual hace parte, otro ejemplo, es cambiar una fecha de cumpleaños y delimitarla a un mes y año de nacimiento o aplicar jerarquías a ciertos valores nominales empleando por ejemplo el carácter **“*”**. Aunque la generalización es una herramienta efectiva para descartar la singularización, no puede garantizar la anonimización de los datos en todos los casos. Los métodos utilizados para la generalización son: **a) agregación y anonimato, b) diversidad y proximidad.**

a Agregación y anonimato

Consiste en generar rangos que puedan agrupar características en macro características o intervalos. Un ejemplo de agregación es reemplazar la edad de un individuo por un rango etario, de manera que la variable “edad” es transformada por intervalos de valor.

Como principales ventajas de esta técnica, se identifica que el incremento de personas o sujetos que comparten un mismo valor de atributo (cuasidentificadores) hace más difícil singularizar a una persona. Como principal riesgo se identifica la posibilidad de inferencia, pues una vez se clarifican las condiciones de la categoría de agregación, es sencillo recuperar el valor inicial que tenía la persona para determinado atributo.

Supresión: consisten en representar los datos de un individuo a través de un símbolo. Un ejemplo de supresión es cuando algunos atributos o variables son reemplazados por un **“*”**. Es decir, se reemplazan todos los valores asignados al atributo **“RELIGIÓN”** por **“*”**. Otro caso, es cuando se opta por eliminar un valor específico del conjunto de datos, por ejemplo, eliminar un valor atípico dentro del registro.

A continuación, se incluye un ejemplo de las técnicas descritas:

Tabla 5. Ejemplo de técnicas

Datos originales			Datos transformados		
Edad	**Departamento**	**Religión**	**Edad**	**Región**	**Religión**
25	Cundinamarca	Católica	20-29	Andina	*
30	Valle del cauca	Cristiana	30-39	Pacífico	*
34	Tolima	Católica	30-39	Andina	*
40	Antioquia	Judía	40-49	Andina	*
36	Nariño	-	30-39	Andina	*
23	Atlántico	-	20-29	Caribe	*
22	Casanare	Católica	20-29	Orinoquia	*

Fuente: Elaboración propia.

Uno de los errores más frecuentes de esta técnica, es no tener en cuenta el número de identificadores como criterio para la anonimización. En el siguiente ejemplo, si se sabe con exactitud que una persona se encuentra en un conjunto de datos, pertenece al rango de edad entre 40-50 años y vive en la ciudad de Bogotá, se sabrá que sufre entonces de colesterol.

Código de municipio	Rango de edad	Enfermedad
11***	30-40	Diabetes
50***	40-50	Colesterol alto
11***	40-50	Colesterol alto
76***	60-70	Diabetes
80***	60-70	Bronquitis

Fuente: Elaboración propia.

b Diversidad y proximidad

El proceso de diversidad y proximidad es similar a la metodología de Agregación. Sin embargo, en este caso los rangos se deben establecer de tal forma que los datos de los individuos puedan ser agrupados y ninguno pertenezca a una categoría individual. En el ejemplo anterior **(ver tabla 4)**, en el rubro Región, hay tres individuos que no comparten grupo con otros participantes, por tanto, según la metodología diversidad y proximidad, se deben encontrar nuevas categorías que puedan agrupar a todos los datos. Es decir que ningún participante puede tener un dato individual.

3.6.3 Seudonimización

La Seudonimización es el proceso mediante el cual se sustituye un dato o atributo por otro, normalmente la información que se intercambia son atributos únicos, que, por sus características, permiten reidentificar a los individuos. Debido al alto riesgo que está asociada con este tipo de datos, la seudonimización no puede garantizar la formulación de registros completamente anónimos, por lo cual no se puede afirmar que sea un método de anonimización. Sin embargo, es una herramienta útil para reducir los riesgos de identificación de los individuos que constituyen los registros (Grupo de trabajo sobre protección de datos del artículo 29, 2014)

La seudonimización emplea cinco técnicas explicadas a continuación;

- **Cifrado con clave secreta**: este es un tipo de encriptación en el que la persona que posee la clave puede de desencriptación puede revertir el proceso.
- **Función Hash:** consiste en elaborar un algoritmo matemático cualquier entrada de tipo numérico o de valor, en una serie de caracteres de una longitud fija. Por ejemplo, el algoritmo permite transformar el nombre Alberto o el nombre Alejandro en códigos de longitud fija de 18 caracteres. Esta función no es reversible como ocurre con el cifrado, pero en caso en que se conozca el rango de los valores de entrada del hash, se podrían pasar los valores por la función con el fin de obtener el valor real de un registro determinado, o se podría buscar el código del hash.

Tabla 6. Ejemplo de función hash

Nombre	Código
Alberto	E7888OP987H987L098
Alejandro	J8787LT9654G672N98

Fuente: Elaboración propia.

- **Función con clave almacenada:** es una técnica de hash que consiste en otorgar una clave secreta a modo de valor de entrada suplementario. Si el atacante no conoce la clave es mucho más difícil identificar el número de combinaciones posible para descifrarla.
- **Función hash con borrado de clave:** consiste en asignar números de manera aleatoria a cada atributo del conjunto de datos y borrar la tabla de correspondencia. Con la aplicación de esta técnica es posible reducir la probabilidad de vinculación entre los datos personales del conjunto de datos y los datos personales del sujeto, contenidos en otro conjunto de datos en el que se ha usado un seudónimo diferente.
- **Descomposición en tokens:** consiste en reemplazar los números de identificación de las personas por números que no son de utilidad para el atacante. Normalmente el número asignado al número de identificación es generado aleatoriamente sin derivar matemáticamente a este último.

Un error frecuente es equiparar las técnicas de seudonimización con las técnicas de anonimización, pues no es suficiente eliminar o reemplazar los atributos para convertir el conjunto de datos en anónimos. La singularización de los sujetos es posible porque quedan identificados como atributos únicos seudonimizado.

Otro de los errores frecuentes en la seudonimización, **i)** es usar la misma clave en varios conjuntos de datos, **ii)** conservar la clave secreta y no guardarla de manera segura o almacenarla junto al conjunto de datos seudonimizado, esto es un riesgo porque el atacante podría llegar a vincular los datos ya seudonimizados con el atributo original, **iii)** usar claves rotatorias para diferentes conjuntos de usuarios y cambiar la clave según el uso, puede facilitar la vinculabilidad de las entradas que corresponden a sujetos determinados.

3.7 Evalúe la utilidad de los datos

El grado de anonimización aplicado a un conjunto de datos afecta la disponibilidad de estos y la usabilidad de la información. La pregunta que surge es si es posible publicar datos útiles y preservar la privacidad de las personas cuya información se encuentra en la base de datos. La relación entre privacidad y utilidad debe ser evaluada a partir de la relación entre el conjunto de datos anonimizado y el cumplimiento de los objetivos de análisis e investigación. Para evaluar la relación entre la utilidad de los datos y el nivel de protección de datos personales se identifican dos procedimientos.

El primero es la aproximación denominada “utility first”, que consiste en anonimizar los datos manteniendo la utilidad de los datos y posteriormente analizar el riesgo de reidentificación. Si se determina que la reidentificación es altamente probable entonces será necesario evaluar, ajustar y rediseñar el proceso de anonimización del conjunto de datos, hasta que el análisis de riesgo sea menor (Gil, 2015, pág. 88).

El segundo procedimiento es el denominado “privacity first”, que ha sido especialmente implementado en el ámbito académico, pero no tan utilizada en la práctica (Gil, 2015, pág. 88). Consiste principalmente en identificar previamente el nivel de privacidad que se desea alcanzar, sin atribuir mayor relevancia al nivel de utilidad de los datos.

Por otra parte, de manera previa al diseño de anonimización, la persona a cargo de las técnicas de anonimización debe identificar previamente cuales son los datos más relevantes para el análisis que se requiere y qué información es de menor relevancia en el análisis. De acuerdo con la relevancia de las variables para la investigación, es necesario tomar decisiones frente a cuáles variables se les hace tratamiento. Por ejemplo, si es importante para la investigación mantener la información de la ocupación y edad de la persona, entonces se deben hacer ajustes en variables de ubicación de residencia u otro tipo de variables del sujeto.

3.8 Documente el proceso de anonimización

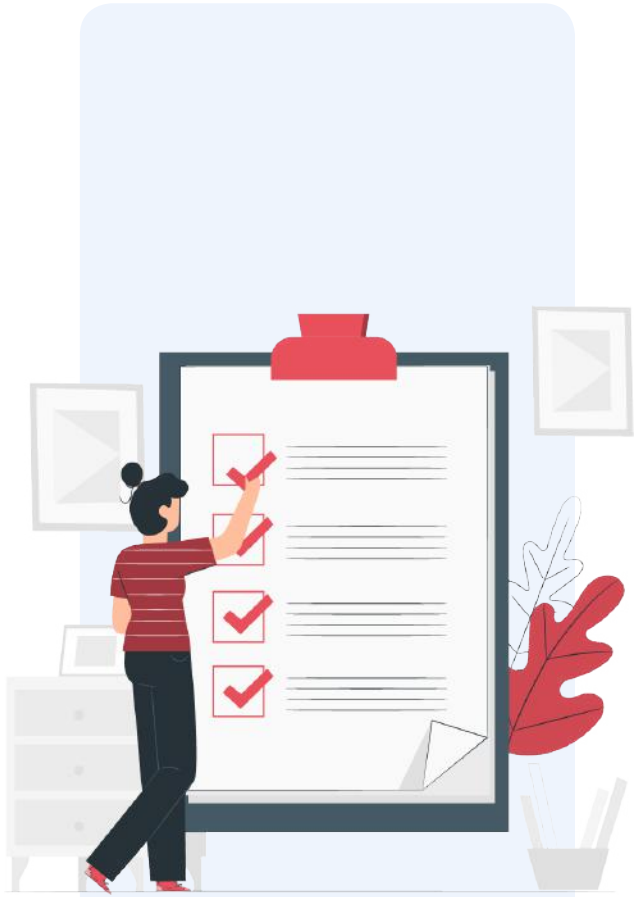
La descripción, características y detalles del proceso de anonimización se deben plasmar en un documento que permita facilitar su revisión, mantenimiento, auditabilidad y replicabilidad. Esta documentación debe estar segura ya que contiene información relevante para la reidentificación de la información. El documento también debe incorporar la descripción de incidentes en caso en que se hayan presentados y las alternativas de solución implementadas para mitigar el impacto o daños causado.

3.9 Publique o comparta la información

Una vez estructurado y aprobado el documento, divulgar y socializar el documento final con los interesados, tanto los participantes el proceso de elaboración como quienes deben ser responsables de su implementación y quienes hacen seguimiento a la aplicación del mismo, con el objeto de garantizar su uso asertivo y procurar a futuro la identificación de mejoras a los lineamientos y el documento mismo como parte de la mejora continua del cumplimiento de requisitos.

4. Gobernanza de datos

Volver al índice



El diagrama metodológico para la anonimización de datos incluye el paso a paso para implementar un proceso estandarizado. Sin embargo, el proceso de anonimización debe estar incluido en un marco de gobernanza de datos propio de la entidad que permita hacer un seguimiento continuo al conjunto de datos publicados y anonimizados. Este marco de gobernanza debe incluir los protocolos y documentación de la metodología de acceso al conjunto de los datos, la identificación de responsables del conjunto de datos, la gestión de claves para garantizar que los datos se mantengan seguros, la revisión periódica de los riesgos de anonimización, la documentación y realización de auditorías a los destinatarios de los datos, la gestión y protocolo en caso de violación de datos personales y la realización de seguimiento a los requisitos de cumplimiento de protección de datos personales.

En la medida en que los procesos de anonimización son necesarios para el desarrollo de proyectos con manejo de datos, se plantea que, entre los principios definidos en las arquitecturas empresariales, las entidades definan un principio relacionado con anonimización de datos por diseño, de acuerdo con el dominio de seguridad o de información del Marco de Referencia de Arquitectura Empresarial y el Modelo de seguridad y privacidad de la información. diseñado por el Ministerio de las Tecnologías de la Información y las Comunicaciones.

Por otra parte, es fundamental contar con un controlador de datos personales dentro de la entidad que esté a cargo de los procesos de anonimización.

A continuación, se incluyen algunos puntos a considerar para el controlador de datos con el fin de que se haga una adecuada gestión y monitoreo de los procesos de anonimización dentro de la entidad.

Tabla 7. Pasos por considerar para una adecuada gestión y monitoreo de procesos de anonimización en la entidad

Tipo	Documentos básicos
Declaración de datos	• Características de los datos, especificación detallada y un ejemplo del conjunto de datos original.
Estado de la anonimización	• Métodos y técnicas aplicadas para anonimizar. • Valoración del riesgo de la técnica de anonimización.
Declaración de datos	• Usuarios que pueden acceder a los datos anonimizados. • Medidas para proteger datos que han sido anonomizados. • Copia del contrato o acuerdo relacionado con la utilización y provisión de los datos.

Fuente: Elaboración propia.

Por otra parte, se recomienda que dentro de la entidad se apliquen medidas de seguridad para prevenir la reidentificación en caso de que los datos anonimizados previamente se filtren o se combinen con otros datos.

- **Medidas de seguridad de tipo gerencial:** se recomienda designar a una persona que se encuentre a cargo de los procesos de anonimización de datos dentro de la entidad. Por otra parte, restringir el intercambio de datos que ya han sido anonimizados y en algunos casos destruir los datos una vez han sido usados para el propósito definido por la entidad. Por último, tener un plan dentro de la entidad para los casos en que se produzca una fuga de información de datos anonimizados. (Office for Government Policy Coordination Korea, 2016)

- **Medidas de seguridad de tipo técnico:** Se recomienda restringir el acceso a archivos de datos que han sido anonimizados, administrar el registro de acceso a la instalación y operación de programas de seguridad. Por último, instalar y operar programas de seguridad para evitar códigos maliciosos. (Office for Government Policy Coordination Korea, 2016).

5. ANEXOS

Volver al índice

5.1 Anexo 1 - Guía para la identificación de riesgos y minimización de reidentificación

A modo ilustrativo se plantea el siguiente ejemplo de un proceso de anonimización de datos donde se identificación los riesgos y se analiza el riesgo de reidentificación. Los datos utilizados fueron creados para ejemplificar los pasos descritos en la sección 3.5 y no corresponden a la realidad.

Ejemplo: en un hospital se hizo una lista de los pacientes que ingresaron por posibles riesgos en el sistema respiratorio. Este ejercicio se llevó acabo como parte de un programa de análisis de datos de la institución para priorizar emergencias y tomar mejores decisiones sobre los pacientes. A pesar de que el ejercicio en un primer momento se pensó a nivel interno, las directivas del hospital decidieron publicar los datos para ayudar a los sistemas estadísticos nacionales de salud y a las investigaciones del sector.

Los encargados del análisis de datos y la consolidación de la información advirtieron a los directivos del hospital que las bases de datos contienen información médica de los pacientes de carácter reservado, por lo que se deben usar procesos de anonimización para disminuir los riesgos de reidentificación de los pacientes evaluados en las listas. A partir de la recomendación el hospital tomo las siguientes acciones para evaluar y disminuir el riesgo de reidentificación su lista:

1 Sistema de publicación del modelo

Para poder definir los riesgos que conllevará la publicación de los datos, el primer paso que se debe llevar acabo es determinar el sistema bajo el cual se va a publicar la información. El hospital previamente había decidido que los datos van a tener un carácter público por los cual los pasos subsiguientes deben tomar las medidas necesarias ante este escenario de publicación.

2 Depuración de identificadores

El segundo paso, dado el carácter público de la lista, es definir cuáles de los datos compilados en la lista son identificadores directos y qué son indirectos. Una vez se hayan definido que datos corresponden a cada categoría, se deben eliminar los identificadores directos. Los identificadores directos son aquellos que permiten reidentificar a una persona instantáneamente.

A modo de ejemplo, las siguientes tablas representan la base de datos del hospital, con información médica y administrativa de 10 pacientes:

#	N° de identificación interno	Documento de identificación	Tipo documento de identificación	Nombre	Sexo	Edad	Fecha de nacimiento	Estado civil	Peso (Kg)	Altura (Cm)	Teléfono	Correo electrónico	¿Tiene acompañante?	Fecha de consulta	Enfermedad	Tratamiento	EPS	ID Médico tratante	Nombre Médico tratante
1	1972	52.638.954	CEDULA	Jairo Olivera	M	69	2/08/1951	CASADO	53	160	312335967	juanchoo@life.com	SI	13/01/2019	Neumonía	Bajo observación	Salud global	1415	Lupe Ramos
2	2715	62.568.357	CEDULA	José Zarco Canabal	M	60	19/03/1960	CASADO	51	162	302405321	jozarco@gnail.com	NO	2/01/2019	Neumonía	Bajo observación	Salud global	1253	Ignacio Escobar
3	4770	8.646.861	CEDULA	Ana Rodríguez	F	51	13/12/1969	CASADO	66	166	301423604	amrodriguezz@hotmail.co	SI	26/01/2019	Bronquitis	Medicación compleja	Vida salud EPS	1398	Alfredo Ruiz
4	3056	101.2358.963	TARJETA DE IDENTIDAD	Mónica Valenzuela	F	14	12/07/2006	SOLTERO	61	165	307517997	monicav@mee.com	SI	15/01/2019	Gripe	Medicación simple	Salud global	1398	Alfredo Ruiz
5	10264	48.726.369	CEDULA	Ximena Castro	F	52	29/07/1968	CASADO	64	168	304579546	xime95@mee.co	SI	27/01/2019	Bronquitis	Medicación compleja	Vida salud EPS	1253	Ignacio Escobar
6	10369	80.967.342	CEDULA	Joaquín Pérez	M	24	24/10/1996	SOLTERO	55	156	302475319	joaquin_perez@yahooo.es	SI	10/01/2019	Gripe	Medicación simple	Vida salud EPS	1398	Alfredo Ruiz
7	10154	52.035.691	CEDULA	Alberto Prado	M	65	19/05/1955	CASADO	57	161	309165850	aprado53@latinmail.co	NO	18/01/2019	Neumonía	Bajo observación	Salud global	1415	Lupe Ramos
8	8043	10.268.936.596	TARJETA DE IDENTIDAD	Juliana Hernández	F	18	27/10/2002	SOLTERO	69	169	300291785	hernandezjuli@yahooo.co	SI	4/01/2019	Gripe	Medicación simple	Salud global	1415	Lupe Ramos
9	5025	9.620.213	CEDULA	Laura Méndez	F	58	8/01/1962	CASADO	62	168	300838673	lauram@gnail.com	SI	28/01/2019	Bronquitis	Medicación compleja	Vida salud EPS	1253	Ignacio Escobar
10	6943	32.598.631	CEDULA	Jaime Bolívar	M	30	9/12/1990	SOLTERO	56	157	309791989	jbolivar_43@gnail.com	NO	2/01/2019	Gripe	Medicación simple	Vida salud EPS	1415	Lupe Ramos

En la tabla anterior, las columnas de datos señaladas en color rojo corresponden a identificadores directos, los cuales no son útiles para un análisis de información y adicionalmente permitirían a un posible atacante reidentificar a los pacientes atendidos, por esta razón deben ser eliminados para la publicación de los documentos. Después de que el hospital reconoció los identificadores directos y los eliminó, la estructura de la información quedó de la siguiente forma:

En la **tabla 3**, el hospital seguirá analizando para determinar los niveles de identificación que se asocian a la información. En los pasos subsiguientes se ejemplifica el proceso que se debe seguir para el cálculo.

Tabla 3. Base de datos del hospital con identificadores indirectos

#	N° de identificación interno	Sexo	Edad	Fecha de nacimiento	Estado civil	Peso (Kg)	Altura (Cm)	¿Tiene acompañante?	Fecha de consulta	Enfermedad o Diagnóstico	Tratamiento	EPS	ID Médico tratante
1	1972	M	69	2/08/1951	CASADO	53	160	SI	13/01/2019	Neumonía	Bajo observación	Salud global	1415
2	2715	M	60	19/03/1960	CASADO	51	162	NO	2/01/2019	Neumonía	Bajo observación	Salud global	1253
3	4770	F	51	13/12/1969	CASADO	66	166	SI	26/01/2019	Bronquitis	Medicación compleja	Vida salud EPS	1398
4	3056	F	14	12/07/2006	SOLTERO	61	165	SI	15/01/2019	Gripe	Medicación simple	Salud global	1398
5	10264	F	52	29/07/1968	CASADO	64	168	SI	27/01/2019	Bronquitis	Medicación compleja	Vida salud EPS	1253
6	10369	M	24	24/10/1996	SOLTERO	55	156	SI	10/01/2019	Gripe	Medicación simple	Vida salud EPS	1398
7	10154	M	65	19/05/1955	CASADO	57	161	NO	18/01/2019	Neumonía	Bajo observación	Salud global	1415
8	8043	F	18	27/10/2002	SOLTERO	69	169	SI	4/01/2019	Gripe	Medicación simple	Salud global	1415
9	5025	F	58	8/01/1962	CASADO	62	168	SI	28/01/2019	Bronquitis	Medicación compleja	Vida salud EPS	1253
10	6943	M	30	9/12/1990	SOLTERO	56	157	NO	2/01/2019	Gripe	Medicación simple	Vida salud EPS	1415

3 Definición del margen de riesgo aceptable

El hospital debe establecer cuál es el nivel máximo de riesgo de reidentificación que aceptará sobre los datos para poder reajustar las tablas de acuerdo con los resultados que obtenga al final del ejercicio.

En estos términos el hospital debe comprender dos cosas: por un lado, el riesgo de reidentificación de un paciente aumenta directamente con el número de identificadores y datos; por otro lado, los riesgos de reidentificación se establecen entre cero y uno, donde cero es mínimo riesgo de identificación y uno es riesgo máximo de identificación. Ningún modelo puede garantizar protección absoluta, por lo cual el hospital debe asumir un riesgo entre cero y uno. Dada la cantidad de los datos el hospital decide asumir un riesgo medio de posibilidad de reidentificación, por tanto, estima que la tabla puede tener un riesgo de 0.7.

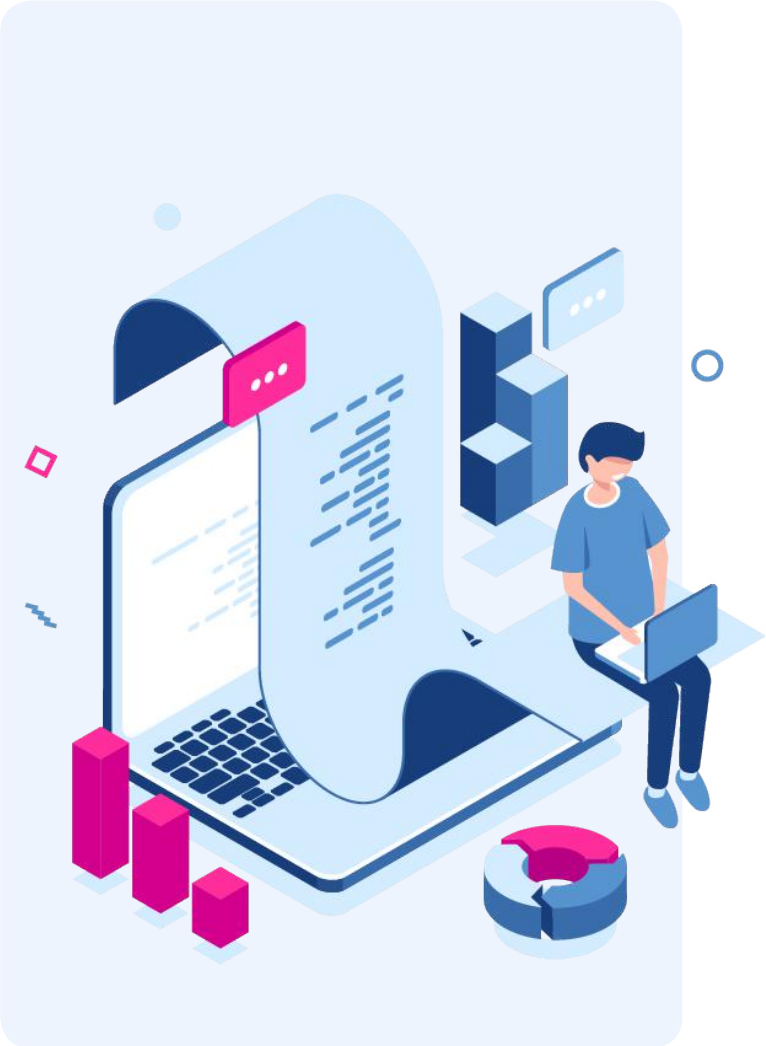
4 Cálculo riesgo de los datos

4.1 Riesgo personal

Como parte del cálculo del riesgo el hospital debe identificar la probabilidad de reidentificación de cada uno de los pacientes dentro de la lista. Para ello se debe hacer la siguiente estimación:

$$\text{Probabilidad de reidentificación} = \frac{1}{\text{(Número de personas con mismos valores)}}$$

La probabilidad de cada individuo está definida por el número de personas que cumplen sus mismas categorías, es decir personas con los mismos cuasi-identificadores. En el caso del hospital, dada la variedad de sus datos, ningún individuo tiene una relación par dentro de la lista y, por tanto la probabilidad de cada paciente es igual a uno. Es decir que individualmente todos los participantes tienen una alta posibilidad de reidentificación igual a 1 dentro de las posibilidades de la lista.



4.2 Método de medición de riesgo adecuado

Teniendo en cuenta que las directivas del hospital decidieron publicar los datos para ayudar a los sistemas estadísticos nacionales de salud y a las investigaciones del sector, estos quedarán accesibles al público en general, por tanto, se debe considerar como riesgo de los datos el valor máximo del riesgo de reidentificación de cada paciente, en este caso el valor sería de 1.

5 Cálculo riesgo del contexto

Para el cálculo de este se debe partir del modelo de publicación de los datos. En el ejercicio planteado, se tiene que los datos serán publicados bajo una política pública de libre acceso por parte de cualquier persona, bajo este supuesto se debe asumir el mayor riesgo posible, es decir, 1.

En caso que se utilice un modelo diferente de publicación como por ejemplo un modelo privado o “no-público” se debe realizar un análisis diferente, en el que se tienen en cuenta aspectos como: posibles ataques a nivel interno de la organización, controles de privacidad y seguridad de la información, reidentificación inadvertida por parte de un analista, filtración de datos y otros.

6 Riesgo total

Una vez se ha calculado el riesgo de la base de datos y el riesgo del contexto, se procede a calcular el riesgo general mediante la multiplicación de estos dos valores, en el caso planteado el riesgo de reidentificación sería 1.

Dado que el riesgo total calculado supera el margen de riesgo aceptable (0.7 definido por el hospital en el paso 3), se deben realizar ajustes a los datos para así disminuir el riesgo de reidentificación.

Como primera medida se debe tener presente el propósito de publicación de la información, en este caso se quiere "publicar los datos para ayudar a los sistemas estadísticos nacionales de salud y a las investigaciones del sector", teniendo en cuenta este objetivo se puede excluir información que no aporte valor para el objetivo de análisis que se pretende alcanzar.

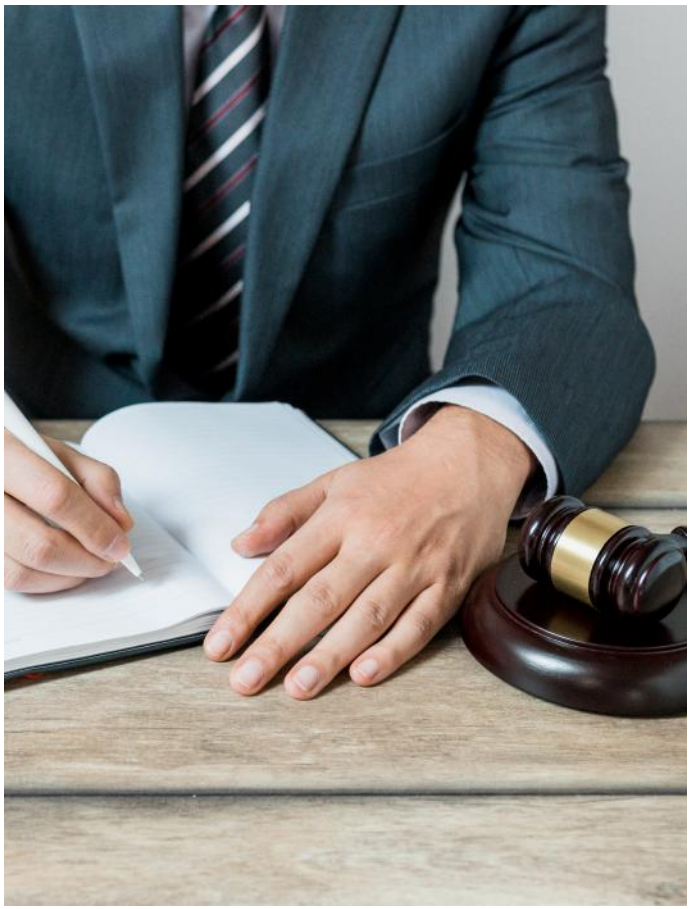
En la base de datos se tiene información médica del paciente al igual que información administrativa. Las variables "Tipo de documento de información", "¿Tiene acompañante?" y "ID Médico Tratante" pueden ser removidas ya que no aportan información desde una perspectiva médica. La variable "ID Médico Tratante" podría ser incluida en un reporte a nivel interno del hospital, en este caso se sugiere utilizar un seudónimo para disminuir el riesgo de reidentificación de los médicos tratantes.

Las variables "Edad" y "Fecha de nacimiento" son redundantes, ya que la primera puede ser calculada a partir de la segunda, a pesar de esto, se opta por conservar la variable "Edad" ya que permite mantener el registro histórico sin importar la fecha de consulta de los datos. Adicionalmente se realiza una generalización de los datos utilizando diferentes rangos etarios, esto permite disminuir el riesgo de reidentificación de los pacientes y a la vez mantener la relevancia médica del registro.

De igual manera se pueden hacer modificaciones de generalización a las variables "Peso", "Altura" y "Fecha de consulta", igual que en el caso anterior, el objetivo principal de estos ajustes consiste en disminuir el riesgo de reidentificación mientras se mantiene la relevancia del registro en el contexto de análisis. Una vez realizados los cambios anteriores se obtiene la siguiente tabla:

#	Sexo	Edad	Fecha de nacimiento	Estado civil	Peso (Kg)	Altura (Cm)	Fecha de consulta	Enfermedad o Diagnóstico	Tratamiento	EPS
1	M	69	2/08/1951	CASADO	53	160	13/01/2019	Neumonía	Bajo observación	Salud global
2	M	60	19/03/1960	CASADO	51	162	2/01/2019	Neumonía	Bajo observación	Salud global
3	F	51	13/12/1969	CASADO	66	166	26/01/2019	Bronquitis	Medicación compleja	Vida salud EPS
4	F	14	12/07/2006	SOLTERO	61	165	15/01/2019	Gripe	Medicación simple	Salud global
5	F	52	29/07/1968	CASADO	64	168	27/01/2019	Bronquitis	Medicación compleja	Vida salud EPS
6	M	24	24/10/1996	SOLTERO	55	156	10/01/2019	Gripe	Medicación simple	Vida salud EPS
7	M	65	19/05/1955	CASADO	57	161	18/01/2019	Neumonía	Bajo observación	Salud global
8	F	18	27/10/2002	SOLTERO	69	169	4/01/2019	Gripe	Medicación simple	Salud global
9	F	58	8/01/1962	CASADO	62	168	28/01/2019	Bronquitis	Medicación compleja	Vida salud EPS
10	M	30	9/12/1990	SOLTERO	56	157	2/01/2019	Gripe	Medicación simple	Vida salud EPS

Teniendo en cuenta la tabla anonimizada se realiza nuevamente el cálculo de la probabilidad de reidentificación para cada paciente, para los pacientes 4,6,8 y 10 se tiene un valor de 0,5 y 0,33 para los demás. Dado lo anterior se tiene como nuevo riesgo de los datos un valor de 0,5.



5.2 Anexo 2 - Marco regulatorio Internacional

A continuación, se mencionan la generalidad de la legislación de la Unión Europea y la de Estados Unidos de América.

- **Unión Europea**

En el año 2016 entró en vigor el Reglamento General de la Protección de Datos (RGPD) y sustituyó la Directiva 95/46/CE que estuvo vigente durante dos décadas. Este reglamento se empezó a aplicar a partir del año 2018. Las normas más estrictas implican: **1)** la protección de datos personales, **2)** las empresas se benefician de igualdad de condiciones. La normatividad se aplica al tratamiento de datos personales fuera del entorno personal, como por ejemplo en el tratamiento de datos en actividades financieras o comerciales.

Mediante el RGDP se estableció un conjunto único de normas directamente aplicables a los Estados miembros, lo cual permite garantizar la libre circulación de los datos de carácter personal entre los Estados miembros y fortalecer la confianza y la seguridad de los consumidores para la consolidación del mercado digital europeo. De acuerdo con el RGPD el tratamiento de datos personales incluye la obtención, registro, organización, estructuración, conservación, adaptación o modificación, extracción, consulta, utilización y comunicación, difusión, supresión o destrucción de datos personales.

Los nuevos derechos que se les otorga a los ciudadanos dentro de este marco regulatorio son **(Comisión Europea, 2018)**:

1. Derecho a recibir información clara y comprensible sobre quien haga el tratamiento de datos.
2. Derecho a solicitar acceso a los datos personales que una organización tenga en su poder.
3. Derecho a solicitar a un proveedor de servicios que transmita sus datos personales a otro proveedor de servicios.
4. Derecho a solicitar que borren los datos personales si ya no se desea que una empresa los trate.
5. Derecho a que una empresa solicite su consentimiento para tratar los datos, informando de manera clara que se hará con los datos personales.
6. Derecho a recibir información por parte de la empresa en caso de pérdida o robo de datos. La empresa causante de la perdida o robo de la información deberá informar y podrá imponer una multa.
7. Mejor protección en línea para los menores de edad, dado que son personas que tienen menos conocimientos de sus derechos y menos conscientes del riesgo al que se ven expuestos.

- **Estados Unidos**

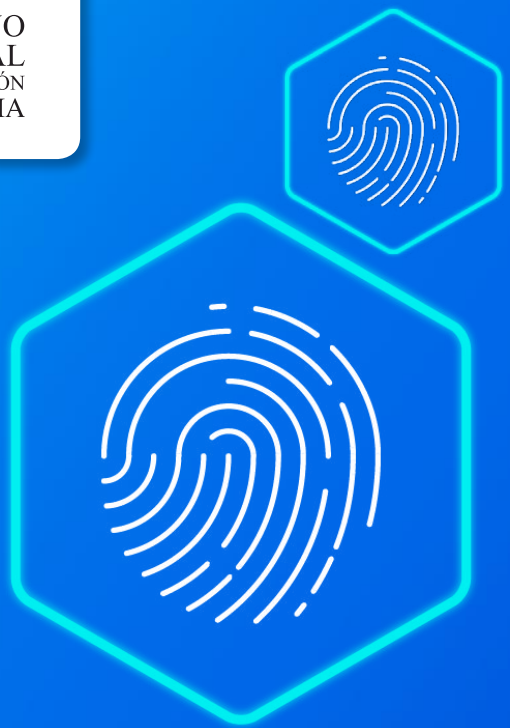
A diferencia de la Unión Europea, la legislación de la protección de datos personales en Estados Unidos no está dictaminada por una única legislación principal. Un conjunto de leyes a nivel nacional y federal permiten emprender acciones de cumplimiento para ello. Algunas de ellas son: La Ley de Transferibilidad y Responsabilidad del Seguro Sanitario (HIPAA, por sus siglas en inglés), la Ley de Protección de Privacidad en Línea para niños (COPPA por sus siglas en ingles) y la Ley de Transacciones de Crédito Justos y Precisos (FACTA por sus siglas en ingles), la Ley de Derechos Educativos y Privacidad de la Familia (FERPA).

Bibliografía

- Agencia Española de Protección de Datos . (2016). Big Data, privacidad y protección de datos. Obtenido de https://www.aepd.es/media/premios/big-data.pdf
- AGESIC. (2017). Unidad Reguladora de control de datos personales. Obtenido de https://www.gub.uy/unidad-reguladora-control-datos-personales/sites/unidad-reguladora-control-datos-personales/files/documentos/publicaciones/Criterios%2Bde%2Bdisociacion%2Bde%2Bdatos%2Bpersonales.pdf
- Comisión Europea Grupo de trabajo sobre protección de datos del artículo 29. (2014). Dictamen 05/2014 sobre técnicas de anonimización. Bélgica. Obtenido de https://www.aepd.es/sites/default/files/2019-12/wp216-es.pdf
- Comisión Europea. (2019). What does data protection ‘by design’ and ‘by default’ mean? Obtenido de https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/obligations/what-does-data-protection-design-and-default-mean_en
- European Parliament. (2015). A comparison between US and EU. Obtenido de http://www.europarl.europa.eu/RegData/etudes/STUD/2015/536459/IPOL_STU(2015)536459_EN.pdf
- Comisión Europea. (2018). Reforma de la protección de datos de la EU. Obtenido de https://ec.europa.eu/commission/sites/beta-political/files/data-protection-factsheet-citizens_es_0.pdf
- Congreso de la República. (2009). Ley 1273 DE 2009. Obtenido de http://www.secretariasenado.gov.co/senado/basedoc/ley_1273_2009.html
- Congreso de la República. (2012). Ley 1581. Bogotá: República de Colombia.
- Congreso de la República. (2012). LEY ESTATUTARIA 1581 DE 2012. Obtenido de http://www.secretariasenado.gov.co/senado/basedoc/ley_1581_2012.html

- Congreso de la República de Colombia. (2008). Ley 1266 de 2008. Obtenido de https://www.habitatbogota.gov.co/transparencia/normatividad/normatividad/ley-1266-2008
- Consejo Nacional de Política Económica y Social - CONPES. (2018). Política nacional de explotación de datos (Big data) - 3920. Bogotá: Departamento Nacional de Planeación, Ministerio de Tecnologías de la Información y las Comunicaciones, Superintendencia de Industria y Comercio.
- Const. (1991). Constitución Política de Colombia.
- DANE. (2018). Guía para la anonimización de bases de datos en el Sistema Estadístico Nacional. Obtenido de https://www.dane.gov.co/files/sen/registros-administrativos/guia-metadatos.pdf
- Data Protection Commission. (2019). Guidance on anonymisation and pseudonymisation. Obtenido de https://www.dataprotection.ie/sites/default/files/uploads/2019-06/190614%20Anonymisation%20and%20Pseudonymisation.pdf
- Decreto 1727 DE 2009. (2009). Obtenido de http://www.suin-juriscol.gov.co/viewDocument.asp?ruta=Decretos/1338429
- Digital Guardian. (Septiembre de 2019). What is HIPAA Compliance? 2019 HIPAA Requirements. Obtenido de https://digitalguardian.com/blog/what-hipaa-compliance
- Emam, K., & Malin, B. (2014). Sharin Clinical Trial Data. Obtenido de Concept and methods for deidentifying Clinical Trial Data.
- GARRIGUES. (2018). ¿Cómo se regula la protección de datos en Latinoamérica y cómo influye el RGPD? Obtenido de https://www.garrigues.com/es_ES/noticia/como-se-regula-la-proteccion-de-datos-en-latinoamerica-y-como-influye-el-rgpd
- Gil, E. (2015). Agencia Española de Protección de Datos. Obtenido de Big data, privacidad y protección de datos: https://www.aepd.es/media/premios/big-data.pdf
- Ministerio de Tecnologías de la Información y las Comunicaciones. (2019). MGGTI.G.GEN.01 – Documento Maestro del Modelo de Gestión y Gobierno de TI.
- MIT. (2007). Spectral anonymization of data. Obtenido de https://dspace.mit.edu/handle/1721.1/42055
- PDPC. (2018). Guide to basic data anonymisation techniques.
- Superintendencia de Industria y Comercio. (2014). Políticas de tratamiento de la información personal. Obtenido de https://www.sic.gov.co/sites/default/files/documentos/Politicas_Habeas_Data_0.pdf

ARCHIVO GENERAL DE LA NACIÓN COLOMBIA



GUÍA DE ANONIMIZACIÓN DE DATOS ESTRUCTURADOS

Conceptos generales y propuesta metodológica

@ArchivoGeneral | Archivo General | YouTube CanalAGNColombia | AGN Colombia

Archivo General de la Nación - Colombia
Establecimiento público adscrito al Ministerio de Cultura
Carrera 6 No. 6-91 - Tel: 328 2888 - Fax: 337 2019
contacto@archivogeneral.gov.co - www.archivogeneral.gov.co
Bogotá D.C - Colombia

La cultura es de todos | Mincultura