



GUÍA BÁSICA PARA LA ANONIMIZACIÓN DE LAS BASES DE DATOS

Código: GU-SI-02

Versión: 01

Fecha de actualización:
31/08/2022

1. OBJETIVO:

Establecer los lineamientos para anonimizar las bases de datos que las Unidades Administrativas utilizarán como insumo para publicar o compartir información de interés con los solicitantes.

2. ALCANCE:

Inicia con el requerimiento de la base de datos o información de la parte interesada, recibido por la Unidad Administrativa y termina con la entrega al solicitante de la base de datos o información anonimizada.

3. RESPONSABLE:

Es responsabilidad del Subdirector de Información y Caracterización y del Profesional Universitario de Información y Caracterización, la adecuada adopción, aplicación y actualización de este instructivo.

4. DEFINICIONES Y TÉRMINOS:

Anonimización de microdatos: proceso técnico que consiste en transformar los datos individuales de las unidades de observación, de tal modo que no sea posible identificar sujetos o características individuales de la fuente de información, preservando así las propiedades estadísticas en los resultados” (Decreto 1743 de 2016: Art. 2.2.3.1.1. Función Pública).

La finalidad de la anonimización es impedir que, a partir de una información o de una combinación de informaciones, se logren identificar sujetos individuales ya sean individuos, empresas o establecimientos, u otro tipo de unidades de observación en un archivo de microdatos (Morales, B. 2017, como se citó en DANE, 2018).

Atributo: también denominado campo de datos, columna de datos o variable. Una información que se puede encontrar en los registros de datos de un conjunto de datos. El nombre, el sexo y la dirección son ejemplos de atributos (ETITC, 2020).

Base de datos: conjunto o colección de datos interrelacionados entre sí, que se utilizan para la obtención de información de acuerdo con el contexto de los mismos y que son almacenados sistemáticamente para su posterior uso” (Sistema de Consultas de Conceptos del DANE, 2018, como se citó en DANE, 2018).

Revisado por: Subdirector(a) de Información y Caracterización	Aprobado por: Director(a) Administrativo de Planeación	Departamento
--	---	--------------

	GUÍA BÁSICA PARA LA ANONIMIZACIÓN DE LAS BASES DE DATOS	Código: GU-SI-02
		Versión: 01
		Fecha de actualización: 31/08/2022

Conjunto de datos: un conjunto de registros de datos. Conceptualmente similar a una tabla en una base de datos relacional típica u hoja de cálculo, que tiene registros (filas) y atributos (columnas) (ETITC, 2020).

Grupos de Interés: individuos u organismos específicos que tienen un interés especial en la gestión y los resultados de las organizaciones públicas. Comprende, entre otros, instancias o espacios de participación ciudadana formales o informales (DAFP, 2019).

Grupos de valor: personas naturales (ciudadanos) o jurídicas (organizaciones públicas o privadas) a quienes van dirigidos los bienes y servicios de una entidad (DAFP, 2019).

Identificación indirecta: la que puede tener lugar como consecuencia de información de una o varias fuentes que por sí misma o en combinación de otros factores puede permitir la reidentificación de las personas cuando sus datos hubieran sido anonimizados. Por ejemplo, la combinación de sexo, edad, lugar de nacimiento y padecimiento de una determinada enfermedad pueden permitir la identificación indirecta de una persona concreta (Agencia Española de Protección de Datos, 2016).

Información estadística: conjunto de resultados y la documentación que los soportan, que se obtienen de las operaciones estadísticas y que describen o expresan características sobre un elemento, fenómeno u objeto de estudio” (DANE, 2017).

Microdato: cada uno de los datos sobre las características de las unidades de estudio de una población (individuos, hogares, establecimientos, entre otras) que se encuentran consolidados en una base de datos” (Decreto 1743 de 2016: Art. 2.2.3.1.1. Función Pública)”.

Proceso estadístico: conjunto sistemático de actividades encaminadas a la producción de estadísticas que comprende, entre otras, la detección de necesidades, el diseño, la recolección, el procesamiento, el análisis y la difusión” (Decreto 1743 de 2016: Art. 2.2.3.1.1. Función Pública).

Reidentificación: identificar a una persona a partir de un conjunto de datos anonimizados. La reidentificación espontánea se refiere a la reidentificación no intencionada debido a un conocimiento especial de las personas (ETITC, 2020).

Registro administrativo: conjunto de datos que contiene la información recogida y conservada por entidades u organizaciones en el cumplimiento de sus funciones o competencias misionales” (Decreto 1743 de 2016: Art. 2.2.3.1.1. Función Pública).

Seudonimización: la técnica de reemplazar un identificador con un valor no relacionado, pero típicamente todavía único (ETITC, 2020).

Operación estadística: aplicación de un proceso estadístico sobre un objeto de estudio que conduce a la producción de información estadística” (Decreto 1743 de 2016: Art. 2.2.3.1.1).

	GUÍA BÁSICA PARA LA ANONIMIZACIÓN DE LAS BASES DE DATOS	Código: GU-SI-02
		Versión: 01
		Fecha de actualización: 31/08/2022

5. CONTENIDO:

5.1 Proceso de anonimización de bases de datos: La finalidad es eliminar o reducir al mínimo los riesgos de reidentificación de los datos, manteniendo la veracidad de los resultados del tratamiento de los mismos, es decir, además de evitar la identificación de los sujetos, los datos anonimizados deben garantizar que cualquier operación o tratamiento que pueda ser realizado con posterioridad a la anonimización no conlleva una distorsión de los datos reales. Un análisis masivo de los datos o macrodatos que puedan derivar de los datos anonimizados no debería diferir del análisis que pudiera obtenerse si se realizara con datos no anonimizados (Agencia Española de Protección de Datos, 2016).

En el proceso de anonimización se deberá producir la ruptura de la cadena de identificación de las personas. Esta cadena se compone de microdatos o datos de identificación directa y de datos de identificación indirecta. Los microdatos permiten la identificación directa de las personas y los datos de identificación indirecta son datos cruzados de la misma o de diferentes fuentes que pueden permitir la reidentificación de las personas, como la información de otras bases de datos del mismo u otro responsable, de las redes sociales, buscadores, blogs, entre otras (Agencia Española de Protección de Datos, 2016).

Los procesos de anonimización se deben de enfocar desde el concepto de protección de datos desde el diseño, lo que significa que los requisitos de privacidad serán tenidos en cuenta desde las etapas iniciales del diseño del sistema de información o del producto utilizado para el proceso de anonimización y durante todo el ciclo de vida de dicho producto o sistema de información (Agencia Española de Protección de Datos, 2016).

Las pautas que se describen en la presente guía pueden servir de base para que las unidades administrativas apliquen el proceso de anonimización aquí sugerido pero no es una estructura conceptual u operacional cerrada, sino que el equipo de trabajo de cada unidad determina si utiliza solo la presente guía o la complementa con otras referencias para emplear el proceso de anonimización en sus bases de datos.

El proceso básico de anonimización del que trata la presente guía consta de tres etapas: **I)** Conformación, capacidades y funciones del equipo de trabajo, **II)** Técnicas básicas de anonimización, y **III)** Metodología sugerida de anonimización.

A continuación se describe cada una de las etapas del proceso básico de anonimización de base de datos:

5.1.1 Etapa I: Conformación, capacidades y funciones del equipo de trabajo.

Cada unidad administrativa deberá contar con un equipo de trabajo que tenga acceso y conocimiento de la base de datos a anonimizar, así como de otros aspectos relacionados con la información de la unidad. Es importante que este equipo de trabajo cuente con las siguientes capacidades:

	GUÍA BÁSICA PARA LA ANONIMIZACIÓN DE LAS BASES DE DATOS	Código: GU-SI-02
		Versión: 01
		Fecha de actualización: 31/08/2022

5.2 Conocer temáticamente el contenido de la base de datos a anonimizar: el fin de contar con el personal especializado en este equipo de trabajo, es apoyar la toma de decisiones de la anonimización a realizar.

5.3 Identificar claramente las operaciones estadísticas y Registros administrativos de la Unidad Administrativa: Antes de realizar la anonimización de una base de datos, el equipo de trabajo deberá verificar si esta es la resultante de alguna operación estadística o de un registro administrativo:

✓ **Caso1: Resultado de operaciones estadísticas:** La entidad tendrá que verificar que la base de datos para la anonimización es la resultante de la finalización de la Fase de Ejecución del Proceso Estadístico. Esto significa que la consistencia de los datos recolectados debe haber sido validada, según: i) las técnicas definidas por la entidad responsable; ii) las variables creadas (o calculadas) a partir de la información recolectada; y, iii) que se encuentren en la base de datos. En caso de valores faltantes, tendrá que verificar que los métodos de imputación hayan sido aplicados.

✓ **Caso 2: Base de datos de registros administrativos:** La entidad debe verificar que la base de datos sea la versión más actual. En este caso, se recomienda que el periodo de tiempo del registro administrativo que se anonimizará, se defina con base en la periodicidad de consolidación de la información. Además, se recomienda revisar la consistencia y calidad de la base de datos teniendo en cuenta la sección Revisión de la consistencia de la base de datos de la Metodología de Diagnóstico de Registros Administrativos.

5.4 Estar al tanto de la normatividad sobre protección de datos e identificación de necesidades de la información: El equipo de trabajo de la unidad administrativa debe realizar una revisión normativa que pueda afectar la publicación de la información sujeta a anonimizar; así mismo, se sugiere identificar las necesidades de información que presentan los usuarios sobre la base de datos. El subproceso se compone de dos pasos:

✓ **Revisión de restricciones de publicación de la información:** Para iniciar esta actividad, el equipo de trabajo tendrá que realizar la revisión de leyes, decretos, resoluciones, convenios institucionales, acuerdos de confidencialidad de la información, estatutos y toda la normatividad que fundamenta el origen del registro administrativo o de la operación estadística de la Unidad Administrativa. El resultado de este análisis determinará si existe alguna norma que impida la publicación de la información de la base de datos a anonimizar.

Un insumo a tener en cuenta en esta actividad para el equipo de trabajo es el alcance de las normas asociadas a la confidencialidad de la información que tiene impacto sobre las entidades; por ejemplo el principio de confidencialidad descrito en la Ley Estatutaria 1266 de 2008 o el artículo 2.2.3.3.5 del Decreto 1743 de 2016, en donde se establecen indicaciones sobre la no exposición de la identificación y ubicación de las unidades de observación al momento de publicar información, A continuación se relacionan algunas normas referentes a la confidencialidad de la información:



GUÍA BÁSICA PARA LA ANONIMIZACIÓN DE LAS BASES DE DATOS

Código: GU-SI-02

Versión: 01

Fecha de actualización:
31/08/2022

- **Ley Estatutaria 1266 de 2008:** El principio de confidencialidad descrito en la ley indica que “Todas las personas naturales o jurídicas que intervengan en la administración de datos personales que no tengan la naturaleza de públicos están obligadas en todo tiempo a garantizar la reserva de la información”.
- **El artículo 2.2.3.3.5 del decreto 1743 de 2016** establece que “las entidades que conforman el SEN ..., deberán guardar la confidencialidad de los datos que permitan la identificación y/o localización espacial de las fuentes, cuando estos fueren recolectados exclusivamente para la producción de las estadísticas oficiales y para fines estadísticos”.
- ✓ **Identificar las necesidades de los usuarios de la información:** Este paso tiene como propósito realizar una caracterización de las necesidades de información de los usuarios sobre la base de datos a anonimizar. En este caso, el equipo de trabajo de la unidad administrativa, debe tener en cuenta las demandas o requerimientos realizados por parte de los usuarios.

Se recomienda elaborar un listado que contenga los requerimientos solicitados, las fechas de solicitud, las variables requeridas, la frecuencia con la que se hacen las solicitudes y las respuestas dadas a dichos requerimientos.

El equipo de trabajo debe hacer una revisión del histórico de las solicitudes, menor a dos años, teniendo en cuenta el volumen de las solicitudes recibidas, la recurrencia y la frecuencia que tiene cada solicitud. Algunas de las solicitudes que podrá revisar el equipo de trabajo se presentan a continuación:

- Solicitudes de información recibidas por distintos usuarios: ciudadanos, academia, empresas, entidades públicas del orden nacional o territorial, organismos internacionales, entre otros.
- Derechos de petición radicados en la entidad durante el último año.
- Necesidades de información propias del área temática de la entidad.
- Otro tipo de requerimientos

Con la información recolectada, el equipo de trabajo podrá clasificar y cuantificar las demandas de información, teniendo en cuenta:

- Tipo de usuarios
- Tipo de solicitudes
- Variables solicitadas
- Nivel de desagregación requerido de la información
- Periodos de la información (tiempo en años) o bases con determinados cortes
 - Frecuencia de las solicitudes
 - Objetivo, finalidad, uso de la información requerida

	GUÍA BÁSICA PARA LA ANONIMIZACIÓN DE LAS BASES DE DATOS	Código: GU-SI-02
		Versión: 01
		Fecha de actualización: 31/08/2022

5.4.1.1 Aplicar las técnicas de anonimización a las bases de datos:

Una ventaja de publicar las bases anonimizadas para la entidad, es la disminución de las cargas operativas para responder solicitudes de información que pueden ser repetitivas o recurrentes.

5.5 Etapa II: Técnicas básicas de anonimización.

En esta etapa se describen las técnicas de anonimización, los casos en los cuales se debe aplicar cada una y la manera correcta de hacerlo, se realizarán ejemplos para una mayor ilustración.

Cabe anotar que para esta etapa se tomó como base, el Manual de anonimización de datos GSI-M-02 versión 1 de diciembre de 2020, propuesto por la Escuela Tecnológica Instituto Técnico Central.

A continuación se relacionan las técnicas de anonimización como son: **1) Supresión de atributos, 2) Supresión de registros, 3) Enmascaramiento de caracteres, 4) Seudonimización, y 5) Generalización.**

5.5.1.1 Supresión de atributos.

La supresión de atributos consiste en la eliminación de una columna de datos o variables, contenidas en una base de datos u hoja de cálculo, esta técnica se puede emplear cuando no se necesita un atributo en el conjunto de datos anonimizados.

Se recomienda aplicar esta técnica al inicio de la anonimización de las bases de datos, puesto que es muy fácil de ejecutar y disminuye la posibilidad de identificación de las unidades de observación de los atributos, como son: Nombre, sexo, dirección, número identificación, entre otras.

Su aplicación debe realizarse eliminando los atributos o si el esquema de la base de datos debe mantenerse para no afectar las características técnicas de la base de datos, se procede a borrar los datos y si es necesario el encabezado. Teniendo en cuenta que es una eliminación real y no un simple ocultamiento de la columna.

Este es el tipo más fuerte de técnica de anonimización, porque no hay forma de recuperar la información de tal atributo.

Ejemplo N° 1: A continuación se presenta un conjunto de datos, el cual contiene las calificaciones de los exámenes de diferentes materias (un examen de un estudiante por cada materia) de los estudiantes del grado 11° de una Institución Educativa. Es de anotar que el solicitante de la información, solo necesita analizar las calificaciones de las pruebas por cada profesor pero sin considerar a los estudiantes ni las materias.

	GUÍA BÁSICA PARA LA ANONIMIZACIÓN DE LAS BASES DE DATOS	Código: GU-SI-02
		Versión: 01
		Fecha de actualización: 31/08/2022

Tabla N° 1. Calificaciones estudiantes grado 11°.

Estudiante	Profesor	Materia	Calificación
Alejandra Zapata G.	Andrés Villareal	Español	5.0
Mario Escobar A.	Andrés Villareal	Español	2.3
Fernanda Restrepo	José A. Jaramillo	Matemáticas	4.7
Carolina Hernandez	José A. Jaramillo	Matemáticas	3.1
Javier Bermudez C.	Tatiana Ríos A.	Química	4.2
Lina M. Castro	Tatiana Ríos A.	Química	3.9
Camilo A. Urrego Z.	Tatiana Ríos A.	Química	4.2

Fuente: Adaptado de (ETITC, 2020).

Para anonimizar la situación planteada en el ejemplo 1, utilizando la técnica de supresión, se debe eliminar los atributos “Estudiante” y “Materia”, quedando la base de datos anonimizada de la siguiente manera:

Tabla N° 2. Calificaciones estudiantes grado 11° (anonimizada).

Profesor	Calificación
Andrés Villareal	5.0
Andrés Villareal	2.3
José A. Jaramillo	4.7
José A. Jaramillo	3.1
Tatiana Ríos A.	4.2
Tatiana Ríos A.	3.9
Tatiana Ríos A.	4.2

Fuente: Adaptado de (ETITC, 2020).

5.5.1.2 Supresión de registros.

Consiste en la eliminación de una fila o registro completo de un conjunto de datos. Esta técnica perturba a diferentes atributos al mismo tiempo, por lo que se debe tener cuidado al momento de suprimir el registro, de manera que no se afecte la información solicitada por el usuario o que la entidad desea publicar.

Se debe utilizar en la eliminación de registros atípicos que son únicos o no cumplen con otros criterios y no se debe tener en el conjunto de datos anonimizados. Esta técnica se puede utilizar antes o después de aplicadas otras técnicas como la de generalización.

Para hacer uso de esta técnica se debe eliminar todo el registro atípico o que contiene información que no es relevante para el solicitante o para la publicación que la entidad desea realizar.

Ejemplo N° 2: Tomando como referencia la base de datos del ejemplo N° 1, un usuario

	GUÍA BÁSICA PARA LA ANONIMIZACIÓN DE LAS BASES DE DATOS	Código: GU-SI-02
		Versión: 01
		Fecha de actualización: 31/08/2022

solicita la información de las calificaciones de los estudiantes para analizar las puntuaciones de las materias de matemáticas y química por cada profesor pero sin considerar a los estudiantes.

Para aplicar la técnica de supresión de registros en este ejemplo, se debe eliminar las filas que contengan la materia “español”, dado que el solicitante no requiere los nombres de los estudiantes y para seguir con un procesos de anonimización más exigente, la entidad debe proceder con la eliminación de la columna “Estudiante”, empleando así la técnica de supresión de atributos, a continuación se observa cómo queda la base de datos anonimizada aplicando dos técnicas.

Tabla N° 3. Calificaciones de matemáticas y química estudiantes grado 11° (anonimizada).

Profesor	Materia	Calificación
José A. Jaramillo	Matemáticas	4.7
José A. Jaramillo	Matemáticas	3.1
Tatiana Ríos A.	Química	4.2
Tatiana Ríos A.	Química	3.9
Tatiana Ríos A.	Química	4.2

Fuente: Adaptado de (ETITC, 2020).

5.5.1.3 Enmascaramiento de caracteres.

Su objetivo es ocultar a partir de símbolos como el asterisco (*) o una equis (X) algunos caracteres de un valor de datos.

Esta técnica debe usarse en los casos donde el valor de los datos está compuesto por una serie de caracteres y al ocultar parte de estos, garantiza la anonimización deseada y dificulta en gran medida la reidentificación.

La forma de empleo de esta técnica se realiza reemplazando los caracteres apropiados con un símbolo que debe escogerse, así mismo se debe establecer un número fijo o variable de caracteres a reemplazar, por ejemplo, para números de tarjetas de crédito se utiliza un número de caracteres fijo y para las direcciones de correo electrónico se utiliza un número de caracteres variables.

Se debe tener en cuenta si la longitud de los datos originales proporciona información sobre los datos originales. El conocimiento de la materia es fundamental, especialmente para el enmascaramiento parcial, a fin de garantizar que los datos correctos estén enmascarados. En cuanto al enmascaramiento completo, el atributo podría suprimirse alternativamente a menos que la longitud de los datos sea de alguna relevancia.

Ejemplo N° 3: A continuación se muestra el número de identificación de los estudiantes de diferentes cursos de una Institución Educativa de Itagüí, quienes aprobaron el año cursado.

	GUÍA BÁSICA PARA LA ANONIMIZACIÓN DE LAS BASES DE DATOS	Código: GU-SI-02
		Versión: 01
		Fecha de actualización: 31/08/2022

Tabla N° 4. Calificaciones promedio de Octavo (8) a once (11) grado.

N° Identificación	Promedio	Curso
1023358625	4.3	10
1023359729	3.9	11
1026789237	4.6	8
1026925823	3.7	9
1027125954	4.0	10

Fuente: Adaptado de (ETITC, 2020).

Luego de aplicar la técnica de enmascaramiento de caracteres a la columna “N° Identificación” de la Tabla N° 4, se obtiene como resultado la tabla N° 5 ya anonimizada:

Tabla N° 5. Calificaciones promedio de Octavo (8) a once (11) grado (anonimizada).

N° Identificación	Promedio	Curso
10XXXXXXXXX	4.3	10
10XXXXXXXXX	3.9	11
10XXXXXXXXX	4.6	8
10XXXXXXXXX	3.7	9
10XXXXXXXXX	4.0	10

Fuente: Adaptado de (ETITC, 2020).

5.5.1.4 Seudonimización (codificación).

Radica en la sustitución de datos de identificación (Nombre, número de cédula, NIT, entre otros) por valores numéricos, alfabéticos o alfanuméricos que deberán inventarse. Laseudonimización también denominada “codificación”, puede ser irreversible o reversible.

Laseudonimización irreversible ocurre cuando los valores originales se eliminan permanentemente, es decir no se hace respaldo de los datos originales, mientras que laseudonimización reversible se presenta cuando el propietario de los valores originales los guarda correctamente y de ser necesario los recupera para vincularlos alseudónimo.

Un mismo individuo puede ser representado con el mismoseudónimo (codificación creada) en diferentes conjuntos de datos, esto se denominaseudónimos persistentes, así mismo, es posible emplear variadosseudónimos para representar a un mismo individuo en diversos conjuntos de datos, esto con el fin de reducir la probabilidad de reidentificación al confrontar los valores con los diferentes conjuntos de datos. Losseudónimos también se pueden generar de forma aleatoria o determinista.

Para utilizar esta técnica, se deben reemplazar los valores de los atributos respectivos con valores creados, por lo tanto, se debe realizar previamente una lista de valores inventados

	GUÍA BÁSICA PARA LA ANONIMIZACIÓN DE LAS BASES DE DATOS	Código: GU-SI-02
		Versión: 01
		Fecha de actualización: 31/08/2022

y escoger aleatoriamente de dicha lista con el fin de sustituir cada uno de los valores originales. Los valores inventados no deben tener relación con los valores originales.

Al asignar seudónimos se recomienda no volver a utilizar seudónimos utilizados anteriormente, sobre todo si fueron originados al azar. También se debe evitar el uso del mismo generador de seudónimos sobre varios atributos.

Ejemplo N° 4: Un ciudadano solicita las calificaciones de los estudiantes que ocuparon el primer lugar al finalizar el año escolar 2021 de una Institución Educativa en el Municipio de Itagüí, con el fin de garantizar la protección de los datos sensibles de los estudiantes se debe aplicar la seudonimización a los nombres de estos.

Tabla N° 5. Calificaciones estudiantes (8° - 11°) que ocuparon primer puesto en 2021.

Estudiante	Grado	Calificación
Sandra María Posada	9	4.1
Óscar Arboleda R.	10	3.9
Carlos Bedoya M.	9	4.7
Manuela Lopez O.	8	4.3
Luis Carlos Benjumea	11	4.5
Laura P. Jaramillo	10	3.7
Juan C. Montes	8	4.9

Fuente: Adaptado de (ETITC, 2020).

Luego de anonimizar el conjunto de datos empleando la técnica de seudonimización, se obtiene como resultado la tabla N° 6:

Tabla N° 6. Calificaciones estudiantes (8° - 11°) que ocuparon primer puesto en 2021 anonimizada.

Estudiante	Grado	Calificación
542015	9	4.1
029463	10	3.9
348213	9	4.7
410078	8	4.3
958647	11	4.5
106925	10	3.7
675213	8	4.9

Fuente: Adaptado de (ETITC, 2020).

	GUÍA BÁSICA PARA LA ANONIMIZACIÓN DE LAS BASES DE DATOS	Código: GU-SI-02
		Versión: 01
		Fecha de actualización: 31/08/2022

5.5.1.5 Generalización.

Busca disminuir la precisión de los datos con el fin de evitar la reidentificación de información sensible a partir del cruce de diferentes bases de datos a los que pueda acceder un tercero, un ejemplo simple de esta técnica, se observa cuando a la entidad le solicitan información de un determinado tema y la base de datos que posee el municipio cuenta con atributos como la edad y la ubicación de personas, por lo tanto al emplear la generalización, lo que se hace es generar un rango de edad y una dirección más general por cada registro relacionado con el atributo correspondiente. Esta técnica también se conoce como recodificación.

La técnica de generalización se recomienda usar para aquellos valores de los datos que pueden recodificarse y el conjunto de datos resultantes, es útil para el solicitante o para el fin (publicación) que lo requiera la organización.

Para implementar esta técnica, el equipo de trabajo de la unidad administrativa deberá crear categorías de datos y reglas apropiadas para traducir datos. Se recomienda que después de aplicar la técnica de la generalización, se eliminen (técnica supresión de registros) aquellos registros que se puedan leer o identificar fácilmente, de manera que el solicitante no pueda llegar a los datos originales. Así mismo se deben diseñar rangos de datos con tamaños adecuados, toda vez que los rangos de datos que son demasiado grandes se pueden "modificar" mucho, mientras que los rangos de datos que son demasiado pequeños se modifican levemente, facilitando la identificación de los valores originales.

Al momento de crear los rangos se debe establecer un valor superior y uno inferior, los cuales se extraen de la información con la cuenta la entidad, un ejemplo típico se encuentra en la estratificación, pues las unidades residenciales del municipio de Itagüí están clasificadas en cuatro 4 estratos (1 Bajo bajo, 2 Bajo, 3 Medio Bajo, 4 medio) entonces el límite inferior será ≥ 1 y el límite superior es ≤ 4 , entre estos valores se definen los rangos propuestos por el equipo de trabajo. A continuación se ilustra un ejemplo de aplicación de la técnica de generalización.

Ejemplo N° 5: Tomando como referencia el ejemplo N° 4, específicamente la tabla N° 6, la cual cuenta con el nombre de los estudiantes (a los cuales se les aplicó previamente la técnica de seudonimización), el curso y las calificaciones de cada uno.

Primero se crea el rango para cada registro relacionado con el atributo "Grado", teniendo en cuenta el límite inferior y superior de los mismos.

	GUÍA BÁSICA PARA LA ANONIMIZACIÓN DE LAS BASES DE DATOS	Código: GU-SI-02
		Versión: 01
		Fecha de actualización: 31/08/2022

Tabla N° 7. Creación de rangos de los grados de una Institución Educativa de Itagüí

Grado original	Rango de Grado creado
Límite Inferior	≥ 6
6	6-7
7	7-8
8	8-9
9	9-10
10	10-11
11	10-11
Límite superior	≤ 11

Fuente: Adaptado de (ETITC, 2020).

Luego de la generalización del grado de la Institución Educativa de Itagüí, Así quedaría el conjunto de datos anonimizado:

Tabla N° 8. Calificaciones estudiantes (8° - 11°) que ocuparon primer puesto en 2021 anonimizada con técnica de generalización.

Estudiante	Grado	Calificación
542015	9-10	4.1
029463	10-11	3.9
348213	9-10	4.7
410078	8-9	4.3
958647	10-11	4.5
106925	10-11	3.7
675213	8-9	4.9

Fuente: Adaptado de (ETITC, 2020).

5.5.2 Metodología sugerida de anonimización.

A continuación se relaciona una serie de pasos recomendamos para que los equipos de trabajo de cada Unidad Administrativa ejecuten la debida anonimización de las bases de datos, dichos pasos fueron adaptados a partir del Manual de anonimización de datos GSI-M-02 versión 1 de diciembre de 2020, propuesto por la Escuela Tecnológica Instituto Técnico Central.

5.5.2.1 Determinar la forma de publicación de la base de datos o conjunto de datos anonimizados.

El equipo de trabajo de la unidad administrativa, deberá definir los canales de comunicación (correo electrónico, correo certificado, página web, redes sociales, entre otros) para publicar o compartir el conjunto de datos anonimizados, para ello deberá establecer si estos son de interés público (cualquier persona) o no público (solo le interesa a la(s) persona(s) o

	GUÍA BÁSICA PARA LA ANONIMIZACIÓN DE LAS BASES DE DATOS	Código: GU-SI-02
		Versión: 01
		Fecha de actualización: 31/08/2022

entidad(es) solicitante(es)).

5.5.2.2 Evaluar la posibilidad de reidentificación de los datos sensibles originales.

El equipo a cargo de la anonimización realizará las pruebas necesarias (un ejemplo es el cruce de bases de datos a las cuales tiene acceso la ciudadanía en general versus la información anonimizada resultante).

5.5.2.3 Clasificar los atributos de los datos.

Los responsables de la anonimización del conjunto de datos, los deberán clasificar como identificadores directos, identificadores indirectos o no identificadores, toda vez que esto incide en la forma como se ejecutarán posteriormente los atributos.

5.5.2.4 Eliminar atributos no utilizados.

El equipo responsable, deberá eliminar permanentemente de la base de datos anonimizada, cualquier atributo que no sea requerido para el objetivo por el cual requirió el solicitante la información, esto con el fin de reducir el riesgo de identificación de los datos originales.

5.5.2.5 Anonimizar identificadores directos e indirectos.

Para anonimizar los identificadores directos e indirectos, se realiza aplicando las diferentes técnicas de anonimización y en cuando sea conveniente se emplean en combinación. Cabe anotar que los registros atípicos de ser posible se deben eliminar, para ello lo debe analizar el equipo de trabajo.

5.5.2.6 Realizar una mayor anonimización de ser necesario.

Una vez ejecutado los 5 pasos anteriores, en caso de que los responsables de la información anonimizada observen una alta posibilidad de reidentificación de los datos originales, se recomienda aplicar más anonimización empleando las técnicas antes descritas y repitiendo los 5 pasos previamente indicados.

5.5.2.7 Evaluar el resultado de la información anonimizada.

El equipo de trabajo deberá analizar la base de datos anonimizada con el fin de evaluar si esta satisface el objetivo buscado por el solicitante o el propuesto por la entidad si es para hacerla pública. De no cumplir con dicho objetivo, se debe replantear el proceso de anonimización ejecutado o se puede determinar si la anonimización para el conjunto de datos de interés es factible o no, toda vez que se pueden presentar la imposibilidad de anonimizar los microdatos. En estos casos los responsables tendrán que analizar la normatividad y procedencia para compartir la información a pesar de los riesgos encontrados.

	GUÍA BÁSICA PARA LA ANONIMIZACIÓN DE LAS BASES DE DATOS	Código: GU-SI-02
		Versión: 01
		Fecha de actualización: 31/08/2022

5.5.2.8 Documentar el proceso de anonimización

El equipo de trabajo definido por la Unidad Administrativa, deberá documentar los pormenores del proceso de anonimización, estándares utilizados y los controles establecidos para futuras aplicaciones.

La importancia de documentar este proceso, les facilitará a los equipos de trabajo el análisis, el sostenimiento, la actualización y las auditorías internas o externas a los que sean sometidos. Es de anotar, que la documentación resultante de este proceso de anonimización debe protegerse y estar segura, puesto que a partir de los estándares definidos en dicho proceso se podría llegar fácilmente a la reidentificación de los datos.

6. DOCUMENTOS DE REFERENCIA

- Departamento Nacional de Estadística – DANE. (2016). Decreto 1743 de 2016, “Por el cual se reglamenta el artículo 160 de la ley 1753 de 2015, se adiciona el título 3 a la parte 2 del libro 2 del Decreto 1170 de 2015 Único del Sector Administrativo de Información Estadística”. Recuperado el 26 de abril de 2022, de DANE – Departamento Nacional de Planeación:
https://www.dane.gov.co/files/sen/normatividad/decreto_1743_noviembre_1_2016.pdf
- Agencia Española de Protección de Datos - AEPD. (2016). *Orientaciones y garantías en los procedimientos de ANONIMIZACIÓN de datos personales*. Recuperado el 6 de abril de 2022: <https://www.aepd.es/sites/default/files/2019-12/guia-orientaciones-procedimientos-anonimizacion.pdf>
- Departamento Administrativo Nacional de Estadística - DANE. (2017). Lineamientos para el Proceso Estadístico en el Sistema Estadístico Nacional. Recuperado el 29 de marzo de 2022, de SEN - Sistema Estadístico Nacional-Colombia:
https://www.dane.gov.co/files/sen/normatividad/Lineamientos_Proceso_Estadistico.pdf
- Departamento Administrativo Nacional de Estadística - DANE. (2018). *Guía para la anonimización de bases de datos*. Recuperado el 30 de marzo de 2022, de SEN - Sistema Estadístico Nacional-Colombia:
https://www.sen.gov.co/files/RegistrosAdministrativos/Gu%C3%ADa_para_la_anonimizaci%C3%B3n_de_bases_de_datos.pdf
- Departamento Administrativo de la Función Pública – DAFP. (2019). *Glosario “Formulario único de Reporte de Avances de la Gestión – FURAG”*. Recuperado el 2 de mayo de 2022: <https://www.funcionpublica.gov.co/documents/28587410/35245587/glosario.pdf/d2ba1d38-d973-7251-06e9-044479090b43?t=1549907117050>

	GUÍA BÁSICA PARA LA ANONIMIZACIÓN DE LAS BASES DE DATOS	Código: GU-SI-02
		Versión: 01
		Fecha de actualización: 31/08/2022

- Departamento Nacional de Planeación – DNP. (2020). *Guía para la definición de diccionario de datos DNP*. Recuperado el 4 de mayo de 2022: <https://colaboracion.dnp.gov.co/CDTI/Oficina%20Informatica/Sistemas%20de%20informaci%C3%B3n/Gu%C3%ADas%20Formatos%20Plantillas/Gu%C3%ADa%20Definici%C3%B3n%20de%20Diccionario%20de%20Datos.pdf?>
- Escuela Tecnológica Instituto Técnico Central - ETITC. (2020). *Manual de anonimización de datos*. Recuperado el 6 de mayo de 2022: <https://www.etitc.edu.co/archives/calidad/GSI-MA-02.pdf>

7. CONTROL DE REGISTROS:

Código	Nombre	Responsable	Lugar de Almacenamiento	Recuperación	Protección	Tiempo de Retención	Disposición Final
No aplica							

8. CONTROL DE CAMBIOS:

Versión	Fecha de Aprobación	Descripción del cambio
01	31/08/2022	Creación del documento

9. ANEXOS:

No aplica