

**UNIVERSIDAD CATÓLICA DE COLOMBIA**  
**FACULTAD DE CIENCIAS ECONÓMICAS Y ADMINISTRATIVAS**

**INFORME FINAL DE PRÁCTICA EMPRESARIAL**  
**PROCESO PARA LA ANONIMIZACIÓN DE MICRODATOS**  
**MANUEL ALBERTO GONZÁLEZ CASTAÑO**

**Bogotá D.C. 2017**



La presente obra está bajo una licencia:  
**Atribución-NoComercial-SinDerivadas 2.5 Colombia (CC BY-NC-ND 2.5)**

Para leer el texto completo de la licencia, visita:  
<http://creativecommons.org/licenses/by-nc-nd/2.5/co/>

**Usted es libre de:**



Compartir - copiar, distribuir, ejecutar y comunicar públicamente la obra

**Bajo las condiciones siguientes:**



**Atribución** — Debe reconocer los créditos de la obra de la manera especificada por el autor o el licenciante (pero no de una manera que sugiera que tiene su apoyo o que apoyan el uso que hace de su obra).



**No Comercial** — No puede utilizar esta obra para fines comerciales.



**Sin Obras Derivadas** — No se puede alterar, transformar o generar una obra derivada a partir de esta obra.

**NOTAS DEL AUTOR**

Este Informe Final de Investigación, en pregrado, se realizó por el estudiante Manuel Alberto González Castaño con código 319262, para cumplir con los requerimientos académicos pertinentes, para su opción de grado, habiendo sido su tutor, la Docente: Tatiana Polanía.

Correspondencia: [magonzalez62@ucatolica.edu.co](mailto:magonzalez62@ucatolica.edu.co); [manuelgon7@hotmail.com](mailto:manuelgon7@hotmail.com).

### **AGRADECIMIENTOS**

A Dios, por su ayuda durante toda mi trayectoria por la universidad, un paso más, dado con el fin de cumplir su propósito.

A mi mamá y hermana, por su apoyo constante y sus oraciones que no han cesado.

A cada profesor que formó mi carácter en estos años y clase tras clase me animaron a mejorar mi desempeño académico y como futuro profesional.

A mis amigos de toda la carrera, en especial a las tres, que siempre me soportaron todas mis quejas, ánimos, desánimos, y que siempre estuvieron ahí cuando lo necesité.

**TABLA DE CONTENIDO**

<b>LISTA DE SIGLAS .....</b>	<b>6</b>
<b>RESUMEN.....</b>	<b>7</b>
<b>ABSTRACT .....</b>	<b>8</b>
<b>PALABRAS CLAVE.....</b>	<b>9</b>
<b>MARCO HISTÓRICO .....</b>	<b>10</b>
<b>MARCO TEÓRICO .....</b>	<b>13</b>
<b>DESCRIPCIÓN DE LA ENTIDAD .....</b>	<b>16</b>
<b>Objeto social .....</b>	<b>16</b>
<b>Misión .....</b>	<b>16</b>
<b>Visión .....</b>	<b>16</b>
<b>OBJETO DE LA PRÁCTICA .....</b>	<b>17</b>
<b>Propósito .....</b>	<b>17</b>
<b>Objetivos .....</b>	<b>17</b>
<b>FUNCIONES DESEMPEÑADAS .....</b>	<b>17</b>
<b>METODOLOGÍA .....</b>	<b>18</b>
<b>Proceso de anonimización.....</b>	<b>18</b>
<b>Viabilidad .....</b>	<b>20</b>
<b>Riesgo .....</b>	<b>21</b>
<b>Ejecución de métodos.....</b>	<b>22</b>
<b>Evaluación .....</b>	<b>22</b>
<b>Documentación y difusión .....</b>	<b>23</b>
<b>RESULTADOS .....</b>	<b>24</b>
<b>CONCLUSIONES .....</b>	<b>25</b>
<b>RECOMENDACIONES .....</b>	<b>26</b>
<b>BIBLIOGRAFÍA .....</b>	<b>27</b>
<b>APENDICES .....</b>	<b>29</b>
<b>Métodos de anonimización .....</b>	<b>29</b>
<b>Métodos basados en la perturbación de datos .....</b>	<b>29</b>
<b>Métodos basados en la no perturbación de datos .....</b>	<b>30</b>

## **LISTA DE SIGLAS**

ANDA: Archivo Nacional de Datos.

DANE: Departamento Administrativo Nacional de Estadística.

DIMCE: Dirección de Difusión, Mercadeo y Cultura Estadística.

DIRPEN: Dirección de Regulación, Planeación, Estandarización y Normalización.

ONE: Oficinas Nacionales de Estadística.

ONU: Organización de las Naciones Unidas

SAS: Statistical Analysis Software.

SEN: Sistema Estadístico Nacional.

## **RESUMEN**

El propósito de este informe es desarrollar de manera clara el proceso de anonimización de datos, desde su viabilidad, hasta su posterior publicación y difusión en el Departamento Administrativo Nacional de Estadística-DANE, a través del Archivo Nacional de Datos- ANDA, en función de responder a las necesidades de la sociedad actual y aprovechar la utilidad de tener información a un alto grado de desagregación. Con el fin de tomar decisiones y analizarlas, entendiendo un poco más el comportamiento económico y social del país. Se realiza una breve descripción de los métodos de anonimización, diferenciando tipos de variables, y su aplicación en algunas encuestas realizadas por el DANE. Finalmente, se describen los resultados de este proceso de anonimización, se realiza la conclusión pertinente y las recomendaciones alusivas a tener presente al momento de realizar el proceso de anonimización de datos.

## **ABSTRACT**

The purpose of this report is to clearly develop the process of data anonymization, from its feasibility, to its subsequent publication and dissemination in Departamento Administrativo Nacional de Estadística-DANE, through Archivo Nacional de Datos-ANDA, in order to respond to the needs of today's society and take advantage of the usefulness of having information to a high degree of disaggregation. With the purpose of make decisions and analyze it, understanding a little more the economic and social behavior, of the country. A brief description of the anonymization methods is made, differentiating types of variables, and their application in some surveys conducted by DANE. Finally, we describe the results of this anonymization process; the relevant conclusion is made and allusive recommendations to keep in mind when performing the process of data anonymization.

## **PALABRAS CLAVE**

Anonimización de microdatos, Archivo, Criterios de Sensibilidad, Datos de carácter personal, Datos estadísticos, Identificaciones indirectas, Información sensible, Microdatos, Operación estadística, Preanonimización de microdatos, Reserva estadística, Riesgo, Seudoanonimización, Variable sensible, Control de revelación estadística.

## MARCO HISTÓRICO

El derecho a la información, ha sido una de las conquistas de las sociedades democráticas, y los gobiernos de los países se enfrentan continuamente a una creciente demanda de datos, generados especialmente por los entes estadísticos, en un entorno diferenciado por la globalización y por la evolución respecto a las nuevas tecnologías de la información y de las comunicaciones.(Sunkel & Trucco, 2010)

La sociedad civil y el Estado demandan información para tomar decisiones acertadas. Por otra parte, existen diversos proveedores públicos y privados que buscan satisfacer las necesidades de los usuarios. El sector público, a través de las agencias estadísticas, es uno de los principales oferentes de información, para la política pública y para la toma de decisiones. Está información es recolectada bajo un compromiso de confidencialidad.(Rivas & Seemann, 2014)

Este principio ha sido la piedra angular de la actividad de los organismos responsables de las estadísticas oficiales. Así lo han reconocido los países, los cuales se han expresado en múltiples ocasiones, a través de las diferentes normativas, comúnmente aceptadas y promulgadas por las Naciones Unidas. En términos prácticos, la confidencialidad no es otra cosa que la promesa que hace quien recolecta la información, al que la suministra, de no revelar datos que afectan su intimidad a un tercero. (ONU, 1984)

En la práctica, se manifiesta una permanente tensión entre dos tendencias existentes: por una parte, existen mayores requerimientos de los usuarios, de más y mejor información, y por otra se reafirma el deber de los organismos estadísticos, de restringir la oferta con el fin de evitar la identificación pública de datos personales, toda vez que la intimidad es un derecho constitucionalmente establecido.

La violación del principio de confidencialidad es posible mediante: la identificación pública de la identidad de una unidad individual, de un atributo de dicha unidad, o la inferencia de la

identidad individual, a partir de alguna información, o mediante la fusión de la información difundida con información de otras fuentes.

Por otra parte, es necesario tener en cuenta el insistente llamado de las organizaciones multilaterales, para que los países hagan los arreglos institucionales necesarios, con el fin de garantizar el acceso a los microdatos, preservando la confidencialidad.

El término “microdatos” hace referencia a los datos de carácter individual, acerca de personas, hogares, negocios u otros entes, que posibilita su distinción específica y que son almacenados, directamente o indirectamente, por los departamentos estadísticos de cada país.(Rytem, 2004)

El Principio de confidencialidad, de los Principios Fundamentales de las Estadísticas Oficiales, promulgado por las Naciones Unidas, señala claramente que: “Los datos individuales recolectados por las agencias estadísticas para compilación, sea que se refieran a personas jurídicas o naturales deben ser estrictamente confidenciales y usados únicamente para propósitos estadísticos”.(ONU, 1984)

La confidencialidad estadística, es garantía de la confianza de la ciudadanía en las estadísticas oficiales. Por ello, el uso de los datos individuales, no con fines estadísticos sino con fines administrativos o de otro orden, que conlleven favorecimientos o daños, no puede ser permitido.(“Revisión Constitucional del Proyecto de Ley Estatutaria número 228 de 2012 Cámara, 156 de 2011 Senado, ‘por medio de la cual se crea la Ley De Transparencia Y del Derecho de Acceso a la Información Pública Nacional,’” 2015)

Sin embargo, en muchos casos, la investigación académica puede añadir valor y reducir el costo de los emolumentos públicos de una operación estadística, al tiempo que puede ser un elemento de gran importancia para el análisis y mejoramiento de la calidad de las estadísticas oficiales. A diferencia del uso para el público en general, donde el responsable del uso no está plenamente identificado y no es posible la exigibilidad de un pacto de confidencialidad definido, el tipo de acceso para la comunidad académica ha sido más flexible, y se ha

potencializado a través de mecanismos como la firma de acuerdos y el licenciamiento del uso. (“Revisión Constitucional del Proyecto de Ley Estatutaria número 228 de 2012 Cámara, 156 de 2011 Senado, ‘por medio de la cual se crea la Ley De Transparencia Y del Derecho de Acceso a la Información Pública Nacional,’” 2015)

Para garantizar la confidencialidad, muchos países han ajustado sus marcos legales. No obstante, éstos, para ser efectivos, deben ser complementados con medidas técnicas y administrativas que garanticen el acceso equitativo a los microdatos. En este contexto se enmarcan los enfoques de control de la identificación pública de datos individuales. Es necesario anotar que nos referimos a datos individuales, bien sea sobre personas naturales o sobre personas jurídicas. (“Departamento Administrativo Nacional de Estadística Manual de Diligenciamiento Muestra Mensual Manufacturera,” 2004)

Dentro de dichos enfoques de control, se encuentran los métodos de restricción del acceso, a través de centros especializados de tratamiento o acuerdos de licencia y los métodos de restricción de los datos propiamente dichos, denominados anonimización, o una combinación de ambos. Por otra parte, los niveles de protección dependen de los tipos de usuarios. El público en general requiere de un nivel de protección distinto del que requiere la comunidad académica. Esta diferenciación ha sido clara en las recomendaciones de las Naciones Unidas. (ONU, 1984)

## MARCO TEÓRICO

El DANE, como principal órgano rector en la investigación y difusión estadística en Colombia, se ha propuesto, de acuerdo a lineamientos internacionales, mejorar la calidad de la información estadística, en especial el acceso a los microdatos, como fuente de valor para la investigación, que son requeridos para el análisis, evaluación y seguimiento de las políticas públicas y así como para el estudio del sector privado y sus intereses.(Bermúdez V, 2015)

El acceso a los microdatos es requerido tanto por el sector privado para mediciones económicas y estándares de la industria, el consumo o el comercio, así como por el sector público para su comparabilidad con otros países y toma de decisiones en políticas tanto económicas como sociales.(Rivas & Seemann, 2014)

La información que se pone a disposición del público contiene un valor agregado que permite su utilización e investigación tanto por personas como empresas para la generación de nueva información,(Gobierno colombiano, 2014) y es por esto, que indica: “los datos suministrados al Departamento Administrativo Nacional de Estadística (DANE), en el desarrollo de los censos y encuestas, no podrán darse a conocer al público ni a las entidades u organismos sociales, ni a las autoridades públicas, sino únicamente en resúmenes numéricos”.(Gobierno colombiano, 1993)

Entonces se crea la necesidad de generar mecanismos que eviten la identificación de quienes proveen la información, dado que la información suministrada puede facilitar la identificación de personas o empresas y permite ser utilizada con fines inapropiados generando desconfianza al entregar sus datos al DANE, y así, afectando considerablemente la credibilidad de la entidad.(DANE, 2014)

En el marco constitucional la legislación incluye la reserva estadística y el habeas data, que propenden por la protección de los datos. La reserva estadística incluye a todos los trabajadores del DANE, naturales y jurídicas, y aplica una cláusula de confidencialidad para todos ellos.

Así, la resolución 1503 de 2011, reglamenta la difusión estadística y conforma un Comité de Aseguramiento de la Reserva Estadística, el cual asesora al director del DANE en la adopción de normas y políticas necesarias para definir los microdatos de uso público.(DANE, 2011b)

Sin embargo, la ley no establece mecanismos o instrumentos para su protección, por lo cual, el DANE como órgano rector, en su compromiso con proteger la privacidad de personas y empresas que le reportan su información, aplicó una nueva estrategia institucional que busca precisamente la confidencialidad de la fuente de información, utilizando para esto, el método de “Anonimización de microdatos” que le permite desarrollar y aplicar técnicas basadas en estándares y buenas prácticas internacionales.(DANE, 2014) Los procedimientos, protocolos y mecanismos de carácter técnico y de seguridad para la anonimización de microdatos. Las condiciones, términos obligaciones y derechos de los usuarios, así como sus registros administrativos de las bases de datos.(DANE, 2011b)

Para los protocolos y mecanismos de carácter técnico y de seguridad, se definen las técnicas de protección de datos estadísticos que tienen por objetivo minimizar las posibilidades de los “intrusos” (usuarios que intentan identificar datos individuales mediante técnicas informáticas y estadísticas) de re-identificar información confidencial. Estas técnicas implican, en algunos casos, pérdida de información detallada, lo que significaría que en ciertas ocasiones no se podrán cubrir las necesidades de determinados usuarios, y este dilema es el que deben enfrentar las ONE a la hora de definir la estrategia de protección de datos. Proteger los microdatos en forma excesiva implica perder calidad en la información. Por el contrario, no tomar ninguna medida para salvaguardar los datos que se difunden generaría gran desconfianza en las ONE, más allá de las implicancias legales que esto conlleva. (Segui Stagno, Federico. 2014).

Es importante también tener en cuenta que no existen técnicas totalmente seguras de protección de los datos, por lo cual, lo mejor es aplicar varias al mismo tiempo. Se debe tener en cuenta la magnitud de estas, puesto que la información a proteger siempre generará alguna pérdida, basándose en el nivel de riesgo aceptable de acuerdo a los lineamientos planteados previamente, en la etapa de riesgos. Este proceso busca un balance entre la protección de los datos y la pérdida de información mediante interacciones, ajustándose, llegando al nivel de riesgo aceptable con el mínimo de pérdida de información. (Segui Stagno, Federico. 2014).

## **DESCRIPCIÓN DE LA ENTIDAD**

### **Objeto social**

Contribuimos al desarrollo del país produciendo y difundiendo información confiable, relevante, oportuna y de calidad.(DANE, 2016)

### **Misión**

Producir y difundir información estadística estratégica para la toma de decisiones en el desarrollo económico y social del país. A partir de su liderazgo técnico ejercer la regulación del Sistema Estadístico Nacional.(DANE, 2016)

### **Visión**

En el 2018 el DANE se consolidará como una institución moderna, innovadora y generadora de conocimiento, y continuará siendo la entidad líder en la producción estadística.(DANE, 2016)

## OBJETO DE LA PRÁCTICA

### Propósito

Aplicar en el Departamento las habilidades, destrezas y conocimientos adquiridos durante toda la carrera profesional, así como desarrollar habilidades investigativas y de manejo de la información, aprendiendo a trabajar en equipo en el medio laboral.

### Objetivos

Desarrollar habilidades personales para el desempeño profesional. Crear y formular procedimientos y técnicas, para facilitar e innovar en los procesos y presentar soluciones reales al DANE, optimizando los recursos que tiene este.

## FUNCIONES DESEMPEÑADAS

- Revisar las metodologías de anonimización para las operaciones estadísticas seleccionadas.
- Preparar las bases de datos de registros administrativos para su aprovechamiento estadístico y aplicación de técnica de anonimización de ser necesario.
- Participar en la vinculación de bases de datos de los registros administrativos para su conformación en registros estadísticos del SEN y su publicación.
- Socializar los avances propuestos y resultados con las áreas técnicas del DANE, grupos de trabajo de la DIRPEN y los responsables de la implementación de mejoras del registro para su uso estadístico.

## METODOLOGÍA

El presente informe de práctica empresarial se realiza a través del método inductivo y su respectivo razonamiento de manera completa, por el cual, la técnica empleada para la recolección de información es observación y descripción del proceso de anonimización de datos en cada uno de sus pasos; se enuncia y estudia el método utilizado para tal fin de acuerdo a investigaciones preliminares, en las cuales permite basarse este informe, y finalmente se dan los resultados de acuerdo al método utilizado.

El principal objetivo en el proceso de anonimización es controlar el riesgo asociado de las personas naturales y jurídicas que brindan su información para fines estadísticos a las ONE. De ahí surge el proceso o la implementación adecuada de la anonimización que busca controlar que dichos microdatos no sean utilizados para fines netamente estadísticos.(DANE, 2014)

Para lograr lo anterior, se debe tener en cuenta que el fin último de este es la utilidad de la información para los usuarios. Así, se procura mantener el aprovechamiento de los datos, intentando introducir el menor ruido posible en los resultados y de igual manera, protegiendo la privacidad de las fuentes de información.(DANE, 2014)

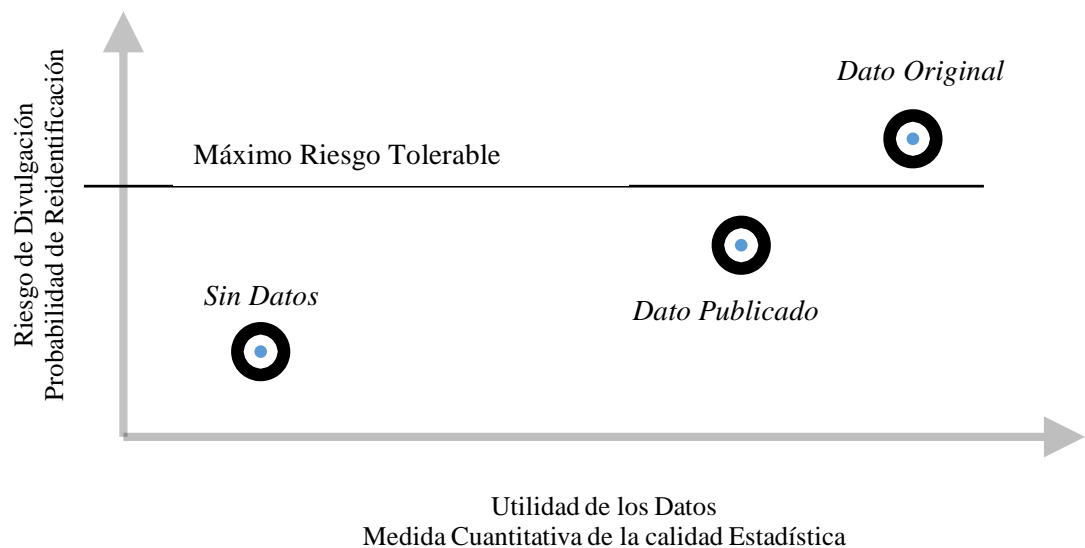
### **Proceso de anonimización**

Para poner a disposición de los usuarios las bases de datos anonimizadas de las operaciones estadísticas del DANE, a través del ANDA, es necesario realizar procedimientos con el fin de no ser detectadas las fuentes de información.(SEN, 2017) Es por ello que el equipo de anonimización de DIRPEN propone distintas técnicas bajo unos parámetros planteados para así, asegurar que la información difundida de los microdatos sea de calidad, optimizando los riesgos y la utilidad de los datos.

El equipo técnico responsable de la operación estadística determina el método de anonimización y los escenarios que minimicen el riesgo a medida que se maximice la utilidad de los microdatos, propuesto por DIRPEN.

El gráfico 1 describe las medidas cuantitativas del riesgo y la utilidad (Duncan et al, 2001).

**Gráfico 1. Mapa de riesgo y utilidad del microdato**



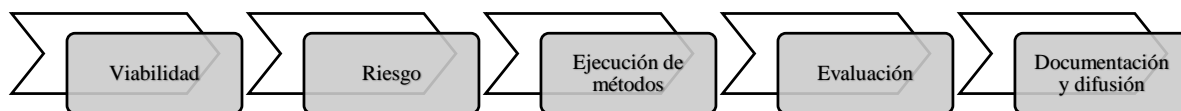
Fuente: Hundepool, Statistical Disclosure Control, 2012

Al observar el gráfico 1 se evidencia la presencia de los posibles riesgos que pueden darse cuando se difunden los datos originales, anonimizados o de no difundir; la línea central muestra el umbral del máximo riesgo tolerable, el cual, al estar por encima, como el caso de los datos originales expone una mayor probabilidad que la fuente sea identificada, aunque esta sea de mejor calidad para su uso. Por debajo del umbral al no publicar información a nivel de microdato, presenta una baja probabilidad de identificación, pero a su vez tiene una baja calidad; En tanto, que el objetivo es presentar los datos anonimizados, por la razón de que existe una baja

probabilidad de detectar al informante y a su vez se preserva la calidad en los datos, asegurando ciertas propiedades estadísticas como el promedio y los totales.

Para abordar el proceso de protección de la confidencialidad se propone basarse en las siguientes 5 etapas:

**Diagrama 1. Etapas del proceso de anonimización**



### **Viabilidad**

En esta etapa se determina la necesidad de proteger la confidencialidad, mediante la interpretación de una normativa que justifique las disposiciones de la privacidad de las fuentes. Se debe analizar las unidades estadísticas y variables.

Se analizan las principales características y el uso de los datos, mediante la observación de la metodología de la encuesta y del formulario de recolección; como resultado se identifica la estructura de los datos y se examina la coherencia y consistencia de los microdatos. Además, se revisa la información sobre las necesidades de los usuarios y se priorizan. (DANE, 2014) Al igual se decide el tipo de publicación a realizar de los microdatos, de acuerdo a las políticas de difusión establecidas y de las necesidades de los usuarios. (DANE, 2011a)

#### ***Establecer la necesidad de la anonimización de la Operación Estadística***

En esta etapa se determina si la fuente autoriza o no su identificación y se define el tiempo en años a partir de los cuales se puede publicar sin riesgo el total de la información, es decir que la información reportada hace (X) años no presenta riesgo y puede ser divulgada.

*Analizar las características principales de la Operación Estadística y el uso de los datos.*

Se identifica la estructura de los datos, como lo es el número de variables, la composición y organización de estos, así como su secuencia.

*Análisis y priorización de las necesidades de los usuarios*

Revisión de las variables de identificación y datos generales. A cada una de estas variables, se le organiza de la siguiente manera:

<b>NO</b>	No se publica
<b>SI</b>	Si se pública sin anonimización
<b>SI-A</b>	Si se pública con anonimización

Se recomienda utilizar el método de redondeo y recodificación de extremos para todas las variables continuas y el método de recodificación para las variables categóricas.

**Riesgo**

Se establecen las necesidades de riesgo de difusión que deben ser protegidos, el cual se realiza mediante una apreciación del método para identificar situaciones donde se presenten inseguridad. Se debe considerar una serie de escenarios potenciales que permitan detectar a determinada fuente y tomar medidas de prevención.

Los escenarios describen la información disponible para el intruso y como este podría utilizar la información para identificar a un individuo o empresa, y define criterios realistas acerca de lo que el hacker puede saber sobre los encuestados. En este paso se determinan los cuasi-identificadores (variables disponibles a los usuarios que permitan la identificación indirecta de las unidades estadísticas) y la forma como se pueden usar dichas variables. Una fuente se encuentra insegura cuando se es capaz de diferenciarla del resto. El riesgo puede ser individual o global.

Una vez establecidos los escenarios de riesgo se adopta el método para estimarlo y se define el umbral, dependiendo del tipo de medida adoptado; estos procedimientos dependen de las características de los datos.

Realizada la evaluación del riesgo se toman medidas para reducirlo a un nivel aceptable, bien sea mediante la aplicación de métodos de anonimización o por el uso controlado de salidas, o, por medio de una combinación entre ambos. Para comparar alternativas es necesario observar el grado de pérdida de la información y el impacto sobre los usos principales de los microdatos.

Se deben considerar los requerimientos de los usuarios, las limitaciones de la estructura de los datos y de acuerdo al método establecido se debe implementar en un software determinado.

### **Ejecución de métodos**

De acuerdo con la evaluación de los riesgos se determina si es necesario realizar la anonimización de los datos.

En esta etapa se realiza la aplicación de los métodos propuestos de anonimización, obteniendo una nueva base de datos con los microdatos protegidos.<sup>1</sup>

### **Evaluación**

Se realiza una auditoría del proceso y un análisis de la pérdida de información; se efectúa principalmente en los métodos de anonimización con perturbación, con el fin de verificar que no se presenta distorsión con los microdatos originales.

Como buena práctica se realizan las rutinas de auditorías, para confirmar que los registros que se encontraban en riesgo ya no lo estén; se analizan si se originan valores inusuales en el procedimiento.

---

<sup>1</sup> Revisar apéndice sobre métodos de anonimización

### **Documentación y difusión**

Es preciso describir los métodos utilizados de protección y de pérdida de información, con el fin de entender lo que ha cambiado o las limitaciones que se pueden presentar debido a la confidencialidad de los datos; además como buena práctica y para procesos futuros en la misma operación estadística y de otras. Este documento debe ser de uso interno, ya que, si el intruso conoce las técnicas realizadas, puede reversar la programación y encontrar determinadas fuentes de información de su interés.

La difusión de los microdatos se realiza por el Archivo Nacional de Datos – ANDA, con su respectiva documentación que facilite la comprensión de los usuarios sobre la encuesta.

## RESULTADOS

Se realiza la anonimización para las series:

- Encuesta Anual Manufacturera.(1995-2015)
- Encuesta Anual de Comercio.
- Encuesta Anual de Servicios.

La aplicación de los métodos de anonimización, da como resultado los microdatos protegidos.

Se desarrollan habilidades y se adquieren conocimientos específicos sobre la utilización de nuevos softwares, como SAS, anonimización de microdatos, además de contribuir al proceso de anonimización.

Los métodos de anonimización usados fueron aplicados de manera exitosa, debido al estudio de Preanonimización en el cual se detallaron e identificaron los tipos de variables, y se estableció para cada encuesta, variables riesgosas, registros y llaves que podrían funcionar como cuasi identificadores o identificadores indirectos, así como unos criterios de riesgos específicos para cada encuesta, permitiendo el diseño y aplicación de los métodos que mejor se adecuaron, de acuerdo a lo anteriormente mencionado.

Las bases de datos fueron modificadas en un porcentaje muy bajo, permitiendo la confiabilidad y representatividad de la muestra, proporcionando la calidad en los datos más que suficiente para su estudio y análisis por parte de usuarios externos.

Se disminuye de manera significativa la probabilidad de riesgo en la identificación o revelación de las fuentes de información, debido a los métodos utilizados para la anonimización de la información.

Se envían las bases a los temáticos de las encuestas para su revisión, aprobación y posterior publicación por la dirección encargada, DIMCE.

## CONCLUSIONES

Los métodos utilizados deben mantener un equilibrio para evitar la pérdida de información, logrando que no se identifiquen las empresas o la población, por lo que se debe tener claridad en el proceso y documentar todo el proceso de análisis de riesgos para que puedan ser revisados. Los usuarios deben ser conscientes de que los datos tienen aplicados métodos de protección.(PERSONALES., 2007)

Cualquier método utilizado se puede especificar, pero el nivel de detalle que se disponga no debe ser suficiente para permitir al usuario recuperar los datos divulgados. Se debe investigar el software y las rutinas disponibles que implementen los métodos de anonimización elegidos con el fin de planear de manera eficiente el proceso de anonimización. Especialmente cuando los métodos perturbativos se aplican a los microdatos, es importante verificar que no haya distorsión o efectos negativos sobre los microdatos originales. Por esta razón, es importante analizar la pérdida de información.

La documentación es una parte esencial de cualquier estrategia de anonimización para una comprensión profunda de los datos, la información sobre las metodologías utilizadas en las diversas etapas de los procesos (toma de muestras, la imputación, validación, etc.) junto con información sobre magnitud de los errores de muestreo, dominios de estimación, etc. son necesarios.

Los microdatos publicados necesitan ir acompañadas por todos los metadatos necesarios. La disponibilidad de programas de lectura para los datos dados a conocer en el software estadístico es importante para los usuarios.

## RECOMENDACIONES

Si los microdatos no están disponibles de inmediato en la web, debido a los procedimientos de acreditación, los metadatos deben estar. Se puede tener en cuenta el evento de poner a disposición de los usuarios en la web un conjunto de microdatos(ONU, 1984) y establecer con exactitud la estructura que tendrá el archivo a ser publicado, con el fin de que el usuario comience a establecer los programas que requiere para ejecutar los datos.

Además de la documentación sobre los métodos de estudio, la documentación sobre el proceso de anonimización es esencial tanto para la auditoría de autoridades externas, así como para garantizar la transparencia hacia los usuarios. El primero puede incluir una descripción de los pasos legales y administrativos de una política de gestión de riesgos junto con la solución técnica aplicada. El último es esencial para que el usuario entienda lo que se ha cambiado o las limitaciones en los datos debido a las restricciones de confidencialidad y las consecuencias de tales acciones.(Dupriez & Boyko, 2010)

Podría ser útil para el usuario, que tenga disponible la lista de variables modificadas por el proceso de anonimización junto con una breve descripción de los cambios introducidos por dicho procedimiento (supresión, modificación, la perturbación, la inserción de los valores que faltan, etc.) o por otras razones (por ejemplo, falta de calidad en los datos originales).

Si un método de reducción de datos se ha aplicado con algunas supresiones, entonces la distribución de tales supresiones se debe dar para una serie de diferentes dimensiones de interés (la distribución por variables, el tamaño del hogar, tipo de hogar, etc.) y cualquier otra estadística que se considere relevante para el usuario.

Si un método de perturbación de datos ha sido aplicado, por razones de transparencia, este debe quedar claro. La información debe proporcionarse en la medida de lo posible, en caso de que las estadísticas se hayan conservado por el método de anonimización o si han sido modificados un poco de orden o de magnitud.

## BIBLIOGRAFÍA

- Bermúdez V, V. E. (2015). FORMULACIÓN DE ESTRATEGIAS DE LIDERAZGO PARTICIPATIVO DEL DANE, PARA UNA EFICIENTE GESTIÓN CON LAS ENTIDADES QUE HACEN PARTE DEL SISTEMA ESTADÍSTICO NACIONAL- SEN, 26. Retrieved from <http://repository.unimilitar.edu.co>
- DANE. (2011a). Código Nacional de Buenas Prácticas para las Estadísticas Oficiales, 16. Retrieved from <http://www.dane.gov.co/index.php/esp/que-es-el-sen/160-uncategorised/4724-codigo-de-buenas-practicas>
- DANE. (2011b). *Resolucion 1503 de 2011, Por el cual se deroga la Resolución No.173 del 2 de abril de 2008 (Por la cual se reglamenta la difusión de la información Estadística), se conforma el Comité de Aseguramiento de la Reserva Estadística y se establecen otras dispos.* Bogotá. Retrieved from <https://www.dane.gov.co/index.php/servicios-al-ciudadano/3796-acceso-a-microdatos-anonimizados>
- DANE. (2014). *Lineamientos para la Anonimización de microdatos*. Retrieved from [http://www.dane.gov.co/files/sen/lineamientos/DSO\\_020\\_LIN\\_08.pdf](http://www.dane.gov.co/files/sen/lineamientos/DSO_020_LIN_08.pdf)
- DANE. (2016). INFORME ANUAL DE SEGUIMIENTO 2015 PLAN INDICATIVO CUATRIENAL 2015 - 2018.
- Departamento Administrativo Nacional de Estadística Manual de Diligenciamiento Muestra Mensual Manufacturera. (2004).
- Dupriez, O., & Boyko, E. (2010). Difusión de archivos de microdatos. Principios , procedimientos y prácticas. *RIEH ( IHSN ) - Red Internacional de Encuestas de Hogares*, 1–76.
- Gobierno colombiano, C. de la R. (1993). Ley 79 de 1993. Retrieved from [http://www.dane.gov.co/files/acerca/Normatividad/Ley79\\_1993.pdf](http://www.dane.gov.co/files/acerca/Normatividad/Ley79_1993.pdf)
- Gobierno colombiano, C. de la R. (2014). LEY 1712 DE 2014, 2014(Marzo 6), 314.
- Hundepool, A., Domingo-Ferrer, J., Franconi, L., Giessing, S., Nordholt, E. S., Spicer, K., & de Wolf, P. P. (2012). *Statistical Disclosure Control. Statistical Disclosure Control*. <http://doi.org/10.1002/9781118348239>
- ONU. (1984). Asamblea General, 9, 9–10. Retrieved from <http://unstats.un.org/unsd/dnss/gp/FP-New-S.pdf>
- PERSONALES., G. D. T. S. P. D. L. P. E. L. Q. R. A. T. D. D. (2007). Dictamen 05/2014 sobre técnicas de anonimización. *Wp216*, 1–24. Retrieved from [http://ec.europa.eu/justice/data-protection/index\\_es.htm%0D](http://ec.europa.eu/justice/data-protection/index_es.htm%0D)
- Revisión Constitucional del Proyecto de Ley Estatutaria número 228 de 2012 Cámara, 156 de 2011 Senado, “por medio de la cual se crea la Ley De Transparencia Y del Derecho de Acceso a la Información Pública Nacional.” (2015).
- Rivas, G., & Seemann, A. (2014). Plataforma integrada de manejo de la información para mejorar las políticas públicas, 109. <http://doi.org/P4, K8, H8>
- Rytem, J. (2004). Manual de organizaciónn estadística: El funcionamiento y organización de una oficina estadística.

SEN. Plan estadístico Nacional 2017-2022 (2017). Retrieved from  
<https://www.dane.gov.co/files/noticias/PEN-2017-2021.pdf>

Sunkel, G., & Trucco, D. (2010). *Nuevas Tecnologías de la información y la comunicación para la educación en América Latina: riesgos y oportunidades*.

## APENDICES

### Métodos de anonimización

Existen varias técnicas que permiten la anonimización de los microdatos. El principal factor para seleccionar una u otra técnica es el soporte internacional que tenga la técnica, esto quiere decir la relevancia del referente internacional en los temas y las experiencias exitosas comprobadas de implementación en otros institutos nacionales de estadística.(DANE, 2014)(Hundepool et al., 2012)

Los métodos se refieren a la limitación de divulgación estadística y pueden clasificarse en dos categorías: métodos basados en la perturbación de datos y métodos basados en la reducción de datos.(DANE, 2014)(Hundepool et al., 2012)

### Métodos basados en la perturbación de datos

Estos se refieren a procedimientos que implican la modificación sistemática de datos (a veces en pequeñas cantidades aleatorias), de manera tal que las cifras no sean lo suficientemente precisas como para revelar información sobre casos individuales. Pueden incluirse nuevos datos, suprimir y/o modificar los existentes beneficiando la confidencialidad estadística.(DANE, 2014)

Las principales técnicas de perturbación de los datos son:

**Tabla 1. Métodos Perturbativos vs Tipo de dato**

<b>Método</b>	<b>Datos Continuos</b>	<b>Datos Categóricos</b>
Enmascaramiento de ruido	X	
Micro-agregación	X	(X)
Cambio de rango(INTERCAMBIO)	X	(X)
Datos aleatorios	X	(X)

Redondeo	X	
Re-muestreo	X	
PRAM		X
MASSC		X

‘X’ indica su aplicabilidad y ‘(X)’ indica su aplicación a datos ordinales categóricos únicamente.

Fuente: Hundepool, Statistical Disclosure Control, 2012

### Métodos basados en la no perturbación de datos

Se refiere a los datos donde al aplicar estas técnicas, no se alteran los datos, sino que se generan supresiones parciales o reducción en el nivel de detalle del dato original. Estas técnicas propenden evitar, o enmascarar los datos únicos, que son atípicos o reconocibles. (Hundepool et al., 2012)

Las principales técnicas de perturbación de los datos son:

**Tabla 2. Métodos No-Perturbativos vs Tipo de dato**

Método	Datos Continuos	Datos Categóricos
Muestreo		X
Recodificación Global	X	X
Codificación superior e inferior	X	X
Supresión local		X

Fuente: Hundepool, Statistical Disclosure Control, 2012

Se debe investigar el software y las rutinas disponibles que implementen los métodos de anonimización elegidos con el fin de planear de manera eficiente el proceso de anonimización. Para el caso de la anonimización de la Encuesta Anual de Comercio, Encuesta Anual de Servicios y Encuesta Anual Manufacturera, se ha elegido el programa Statistical Analysis Software – SAS, en el cual se han desarrollado los programas para el análisis del riesgo y para la aplicación de los métodos de anonimización.