



# Marco ético de los datos

**Piedad Urdinola Contreras**

Directora

**Leonardo Trujillo Oyola**

Subdirector

**Edna Patricia Rangel Barragán**

Secretaria general

**Julieth Alejandra Solano Villa**

Directora técnica

Dirección de Regulación, Planeación,  
Estandarización y Normalización

**Javier Sebastián Ruiz Santacruz**

Director técnico

Dirección de Censos y Demografía

**Juan Pablo Cardoso Torres**

Director técnico

Dirección de Síntesis y Cuentas Nacionales

**Sandra Liliana Moreno Mayorga**

Directora técnica

Dirección de Geoestadística

**Andrea Ramírez Pisco**

Directora Técnica

Dirección de Metodología y Producción  
Estadística

**Andrés Felipe Ortiz Rico**

Director técnico

Dirección de Recolección y Acopio

**Álvaro Enrique Duque Soto**

Director

Dirección de Difusión y Cultura Estadística

**Departamento  
Administrativo  
Nacional - DANE**

## Elaborado por:

### **Sistema de Ética Estadística del DANE (SETE)**

#### **Wilson Herrera Romero**

Profesor titular y miembro Centro de formación en ética y ciudadanía Phronimos de la Universidad del Rosario

#### **David Hernández Zambrano**

Profesor e investigador, miembro del Centro de formación en ética y ciudadanía Phronimos de la Universidad del Rosario

#### **Elizabeth Moreno Barbosa**

Asesora de la Dirección de Regulación, Planeación, Estandarización y Normalización

#### **Orlando Uribe**

Coordinador académico de Ética Psicológica de la Asociación Colombiana de Facultades de Psicología, el Colegio Colombiano de Psicólogos y la Universidad del Rosario

## Revisión y validación:

### **Comité de Administración de Datos**

#### **Viviana Rocío Vanegas Barrero**

Directora  
Dirección de Desarrollo Digital  
Departamento Nacional de Planeación

#### **Cielo Ángela Peña Rodríguez**

Superintendente delegada para la Protección de Datos Personales  
Superintendencia de Industria y Comercio

#### **Cesar Augusto Cruz Aya**

Director de Gobierno Digital  
Ministerio de Tecnologías de la Información y las Comunicaciones de Colombia

#### **Leonardo Trujillo Oyola**

Subdirector  
Departamento Administrativo Nacional de Estadística

#### **Julieth Alejandra Solano Villa**

Directora técnica  
Dirección de Regulación, Planeación, Estandarización y Normalización  
Departamento Administrativo Nacional de Estadística

#### **Luis Martín Barrera Pino**

Jefe  
Oficina de Sistemas  
Departamento Administrativo Nacional de Estadística

### **Consejo Asesor Técnico del Sistema Estadístico Nacional 2022**

Sala Especializada de Modernización Tecnológica del Consejo Asesor Técnico del Sistema Estadístico Nacional (CASEN)  
León Darío Parra Bernal  
Valérie Gauthier Umaña

### **Comité Estadístico Sectorial Infraestructura Tecnológica Mesa Estadística Sectorial TIC**

## Corrección de Estilo

Sonia Marcela Naranjo Morales

## Diagramación

Jazmin Carolina Waltero Argüello

© DANE, 2023

Prohibida la reproducción total o parcial sin permiso o autorización del Departamento Administrativo Nacional de Estadística, Colombia

---

# Tabla de contenido

---

**Introducción**

**5**

---

**1. La gestión ética de los datos**

**6**

1.1. El modelo de Davis

1.2. El modelo del Banco Interamericano de Desarrollo (BID) y el ciclo de vida de los datos

1.3. Síntesis de los dos modelos

1.4. Aplicación del presente modelo en usos no estadísticos

---

**2. Principios éticos de la gestión ética de los datos**

**13**

2.1. Privacidad y seguridad

2.2. Derechos humanos

2.3. No maleficencia

2.4. Confiabilidad: exactitud y precisión

2.5. Relevancia

2.6. Eficiencia y usabilidad o buen uso

2.7. Transparencia

---

**3. Diseño institucional de la evaluación de la gestión ética de los datos del SEN**

**23**

3.1. Lineamientos para la selección

3.2. Componentes

3.3. Proceso

3.4. Pasos por seguir

---

**Anexos A. Descripción de las fases del ciclo de vida de los datos**

**30**

---

**Bibliografía**

**33**

# Introducción

Una de las funciones centrales del Comité de Administración de Datos (CAD), instancia que integra la coordinación del SEN, es “definir el marco ético que orienta[rá] la revisión de los casos, promoviendo el mayor uso y reúso de los datos garantizando que este sea adecuado, responsable y seguro” (Resolución 2259 de 2023 del DANE, art. 3, numeral 3). Esta función del CAD en principio apunta a que su tarea de promover el intercambio seguro de datos dentro de las entidades del SEN se haga de manera ética y responsable. Para este efecto, se entiende que el CAD debe asegurar que las entidades que hacen parte del SEN tengan un proceso de gestión de datos regido por orientaciones éticas que sean acordes con los principios de un Estado Social de Derecho y con experiencias internacionales en relación con el manejo ético de los datos. Además, “el CAD tiene como objetivo, garantizar la implementación de un marco ético para el uso adecuado de los datos” (Ley 2335 de 2023, art. 14, numeral 4) por lo que dicho marco debe cumplir funciones orientadoras para el uso de los datos y presentar los lineamientos para su implementación.

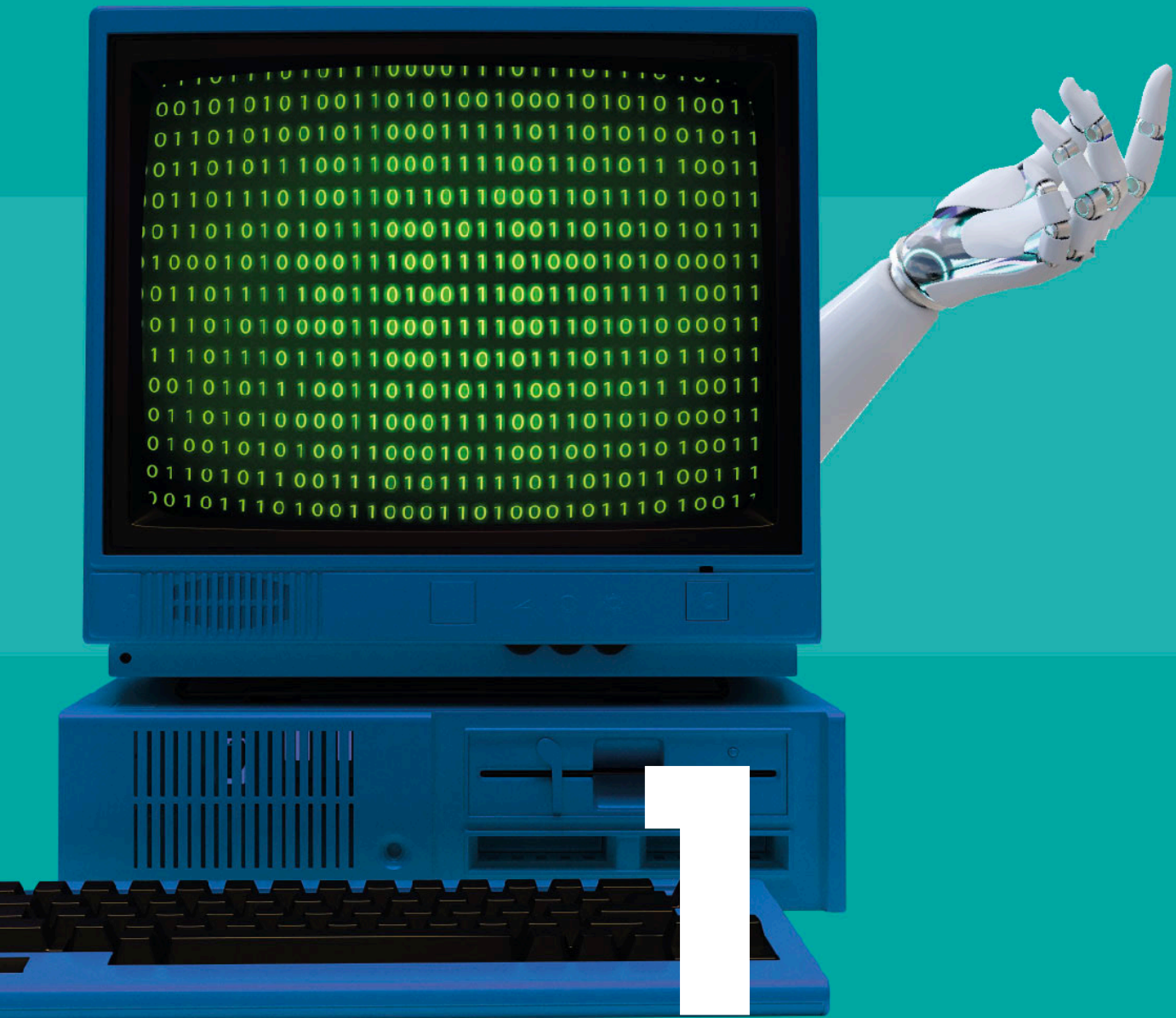
Por su parte, el Sistema de Ética Estadística del DANE (SETE) tiene como uno de sus propósitos a largo plazo la elaboración de un marco ético para el Sistema Estadístico Nacional (SEN). La experiencia del SETE durante el período 2021 - 2022 le permitió

evaluar diecisiete operaciones estadísticas, realizar cuatro seminarios web, escribir varios documentos en torno a temas éticos sobre la producción estadística, desarrollar talleres de sensibilización sobre los estándares éticos que deben cumplir las operaciones estadísticas y diseñar un curso virtual sobre ética estadística. Todas estas actividades le han dado al SETE y, en especial al grupo de trabajo que tiene la función de ejercer la secretaría técnica del Comité de Ética Estadística y de la Sala de apelaciones, la madurez suficiente para enfrentar el reto de elaborar, junto con la secretaría técnica del CAD, una propuesta del marco ético de los datos para el SEN.

En líneas generales, este documento tiene dos objetivos: proponer el conjunto de principios éticos que debe guiar el manejo de los datos en las entidades que hacen parte del SEN y que sirvan de referencia al CAD para garantizar que el intercambio de datos se haga de manera ética y responsable y proponer un diseño institucional para garantizar un alineamiento entre los valores establecidos en el marco ético y las prácticas en el manejo de los datos de las entidades que hacen parte del SEN.

Para la constitución de este marco, el grupo base del SETE realizó la revisión de la literatura sobre el tema y de algunas experiencias internacionales. Además, se llevó a cabo una revisión y una armonización del documento en noviembre de 2023 acorde con lo dispuesto en la Ley 2335 de 2023 “Por la cual se expiden disposiciones sobre las estadísticas oficiales en el país”, sancionada el 3 de octubre de 2023.

El documento se divide en tres partes: en la primera parte se explica cómo se entiende la gestión ética de los datos; en la segunda parte se presentan los valores que configuran el marco ético, y en la última se propone un esbozo del diseño institucional para hacer seguimiento a la aplicación del marco ético en las entidades del SEN.



# La gestión ética de los datos

Establecer el esquema de gobernanza y administración de datos es una de las funciones que, como autoridad estadística, posee el DANE y en ella se contempla la formulación del “marco ético que permita articular la información estadística con el ciclo de las políticas públicas” (Ley 2335, art. 7, numeral 11), por lo que se hace necesario que los planteamientos del presente marco atiendan a las necesidades y las realidades del manejo de la información en el contexto contemporáneo.

Hoy en día, los avances en las tecnologías de la información han generado una serie de dilemas éticos, tanto para la sociedad en general como para las organizaciones, que no son fáciles de resolver. Ello se debe, en parte, a que los principios éticos que deben guiar la recolección, el análisis, la difusión y el uso de los datos son generales y abstractos y las realidades donde se dan estos procesos son complejas y llenas de incertidumbres debido a la diversidad de actores que intervienen y donde cada uno tiene intereses distintos que pueden llegar a ser contrapuestos.

En este sentido, la gestión ética de los datos facilita la toma de decisiones sobre estos dilemas a las organizaciones. De acuerdo con Payam Moula y Per Sandin, la gestión ética puede ser vista como una herramienta que “puede[n] ayudar a los miembros de una organización o una comunidad a mejorar la calidad de sus deliberaciones con el fin de alcanzar decisiones y juicios éticamente informados” (2017, p. 126).

Otro propósito de la gestión ética en las organizaciones consiste en que las decisiones y las acciones relacionadas con el manejo de los datos correspondan con los valores y los principios con las que están comprometidas. Como bien lo señala Kord Davis, las organizaciones deben tratar de “alinearse sus acciones con sus valores” (2012, p. 64). Esta afirmación de Davis apunta a un problema central de la ética de las organizaciones que es evidente en el caso del manejo de los datos: mientras los valores son generales, las decisiones de las organizaciones tienen que ser concretas y particulares para ejecutarse. Aunque las organizaciones, por su propia naturaleza,

tratan de tener procesos estandarizados que faciliten la ejecución de sus acciones, cuando se está frente a situaciones donde hay asuntos éticos relevantes y en las que los principios éticos pueden estar enfrentados, no hay recetas o fórmulas que establezcan de antemano cómo se pueden concretizar dichos principios en acciones. Este reto requiere no solo de diálogo entre los distintos grupos involucrados en el manejo de los datos, sino también de una articulación coherente entre las distintas organizaciones en el proceso de gestión de datos.

Ahora bien, es importante tener en cuenta que el marco ético que aquí se propone se refiere tanto a la gestión de los datos en general como de aquellos que tienen un uso estadístico. En la Ley General sobre las Estadísticas Oficiales elaborada por la Comisión Económica para América Latina y el Caribe (CEPAL), su artículo primero establece que las estadísticas oficiales pretenden “describir de manera representativa los fenómenos económicos, demográficos, sociales y ambientales de un [país o territorio]”. Dicha ley es coherente con la definición presente en la Ley 2033 de 2023 donde se define que “Las estadísticas oficiales son las que permiten conocer la situación económica, demográfica, ambiental, social y cultural de acuerdo con el nivel de desagregación territorial de la operación estadística, y sirven como insumo para la toma de decisiones públicas y privadas, en especial, para la generación, el diseño y el seguimiento de las políticas públicas” (Art. 5, numeral 1). Esta definición recoge lo que usualmente se entiende como estadística: recoger y analizar información preferencialmente cuantitativa para, primero, describir y explicar un fenómeno y, segundo, a partir de allí hacer inferencias y predicciones. Además, la estadística, especialmente la que tiene un carácter oficial, usa toda esta información para contribuir con la formulación y el seguimiento de políticas públicas y el proceso de toma de decisiones.

En este orden de ideas, con este marco se busca que los datos que hacen parte del SEN y que tengan un fin estadístico se recojan, analicen y divulguen cumpliendo con los principios aquí establecidos para así garantizar

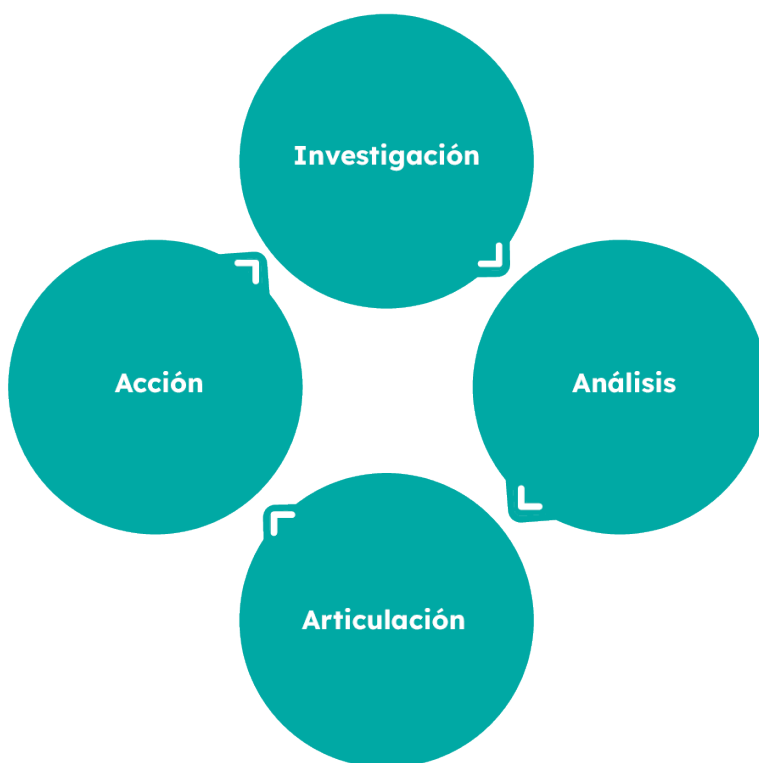
el seguimiento de las disposiciones legales en Colombia. Al mismo tiempo, este propósito puede entenderse como un medio para un objetivo más amplio que es el de contribuir con la construcción una cultura estadística orientada éticamente. Mediante esta construcción se espera que, tanto la ciudadanía general como las entidades públicas y privadas, cuenten con información confiable y apropiada y tengan la capacidad de hacer un buen uso de ella, sobre todo en el diseño y la evaluación de políticas públicas.

Para los efectos de este documento, se han considerado dos concepciones sobre la gestión ética de los datos que son complementarias: una formulada por Davis y la otra por el Banco Interamericano de Desarrollo (BID). A continuación, se describen brevemente cada una de ellas.

### 1.1. El modelo de Davis

La propuesta de Davis apunta a crear dentro de una organización un ecosistema en el manejo de datos que pueda verse retroalimentado constantemente por los valores organizaciones. En este sentido, las distintas actividades están estrechamente encañadas, por lo que los procesos de la organización tienen en todo momento como referente sus valores y al mismo tiempo continuamente se reflexiona sobre cómo los valores pueden aterrizar a las actividades sobre las que ocurren los procesos. El modelo de Davis tiene cuatro componentes: “investigación, análisis, articulación y acción” (2012, p. 27).

**Gráfica 1.** Modelo de Davis



Fuente: Davis, Kord, Ethics of big data, p. 67.



**Investigación:** Davis se refiere fundamentalmente al establecimiento de los valores éticos con los cuales se compromete la organización. En un estado democrático lo ideal es que en las entidades públicas estos valores se establezcan a través de un proceso dialógico en el que participen los grupos de interés más relevantes. Al mismo tiempo, estos valores deben ser compatibles con los principios propios de un Estado social de derecho y su marco jurídico.

**Análisis e investigación:** en este componente se hace una descripción de las prácticas que la organización sigue en el manejo de los datos y se hace una evaluación de en “qué medida estas prácticas se alinean con los valores de la organización” (2012, p. 68).

**Articulación:** teniendo en cuenta los resultados de los componentes de investigación y análisis, aquí se determina la magnitud de las brechas entre los valores y las prácticas y se establecen posibles cursos de acción para superar dichas brechas. Esta parte culmina con un plan de acción que debe ser acordado entre las partes involucradas en cada uno de los procesos en el manejo de los datos.

**Acción:** este último componente trata de la puesta en marcha del plan de acción y su implementación. Para ello, es necesario contar con las capacidades y los recursos necesarios para cumplir con dicho plan. A su vez, en esta etapa se debe contar con indicadores que permitan establecer en qué grado se están acortando las brechas.

## 1.2. El modelo del Banco Interamericano de Desarrollo (BID) y el ciclo de vida de los datos

En su estudio para el BID, sobre la gestión ética de los datos, Cesar Buenadicha, Gemma Galdon, María Paz Herмосilla, Daniel Loewe y Cristina Pombo plantean que uno de los propósitos de la gestión ética de los datos es brindar una herramienta para que las entidades públicas puedan sortear con éxito los riesgos éticos involucrados en el

manejo de los datos, dentro de los cuales están: la violación a la privacidad, la discriminación algorítmica y la opacidad (2019, p. 39). Vale la pena señalar que, a diferencia del modelo de Davis, estos autores parten de los riesgos para establecer los criterios que deben guiar la gestión en sus distintas fases. En ese mismo documento se proponen los siguientes criterios que, dicho sea de paso, fueron tenidos en cuenta para este estudio: la creación del valor público; la identificación de las personas beneficiadas por los datos; el diagnóstico de los datos; la privacidad; la identificación de buenas prácticas, y la transparencia y la rendición de cuentas.

Para estos autores la gestión ética de los datos es un conjunto de procesos “que se debe realizar a lo largo del ciclo de vida de los datos” (2019, p. 40). Este ciclo tiene cinco etapas relevantes e incluye algunas de las etapas contempladas en los modelos de la actividad genérica para las organizaciones estadísticas (GAMSO por sus siglas en inglés) y del proceso genérico del modelo estadístico (GSBPM por sus siglas en inglés) que se implementan en el DANE. Estas etapas son: recolección, almacenamiento, análisis, compartición y archivo.

- **Recolección:** como su nombre lo indica, esta etapa se refiere a “la obtención de los datos por parte de la entidad a través de formas automatizadas o manuales” (2019, p. 40), con fuentes directas o con datos administrativos.

- **Almacenamiento:** en esta fase “se mantienen los datos protegidos de manera segura y se establece la manera cómo se accede a ellos” (2019, p. 40)”.

- **Análisis:** en esta parte “se hace una evaluación de los datos” (2019, p. 40) y se determina la información que sea necesaria para la toma de decisiones.

- **Compartición:** en esta etapa se comparan los datos tanto al interior de la entidad como con entidades externas, que pueden ser públicas o privadas, y con el público en general; en esta etapa se consideran los mecanismos de acceso y difusión.

• **Archivo:** en esta fase se establece cuáles datos se eliminan y cuáles se mantienen. Esto último se hace garantizando la seguridad de los datos y cerraría el ciclo de vida de los datos.

Cabe aclarar que en el contexto colombiano el *Plan Nacional de Infraestructura de Datos* también propone su propio modelo de ciclo de vida de los datos. En este documento el ciclo se determina con las etapas de “creación y recepción, procesamiento, almacenamiento, intercambio, transferencia y compartición, uso y análisis de datos, archivo y preservación (incluyendo la posibilidad de destruir los conjuntos de datos)” (MinTIC et al., 2017 p. 17) las cuales, como puede observarse, coinciden en gran medida con las propuestas por el BID. La principal diferencia es la consideración explícita del procesamiento como una etapa intermedia entre la generación del dato y su almacenamiento; en este caso se reconoce que una vez establecidos los mecanismos y los diseños para la adquisición de los datos

existirá un proceso para su almacenamiento que implicará su transformación. Así, en esta etapa del ciclo de vida también será importante tener una consideración por las implicaciones y el respeto de los principios éticos acá presentados<sup>1</sup>.

### 1.3. Síntesis de los dos modelos

Los dos modelos expuestos no son opuestos sino complementarios. En el primero se establece los momentos que se deben tener en cuenta para alinear los valores con los procesos y, en el segundo, se señalan las fases del ciclo de vida de los datos. En este sentido, se puede decir que, una vez establecidos los principios del marco ético, los componentes de análisis, articulación y acción se pueden aplicar a las cinco fases mencionadas. Para el caso de la propuesta que aquí se hace, los principios que se aplicarían son los que se exponen en el siguiente apartado del documento.

---

<sup>1</sup> Para profundizar más sobre el ciclo de vida de los datos consultar el Anexo A. Descripción de las fases del ciclo de vida de los datos, del presente documento.

**Tabla 1.** Estapas y componenetes del ciclo de vida de los datos

Etapa/componente	Análisis o evaluación	Articulación o plan de acción	Acción o puesta en marcha del plan
Recolección	Aplicación de los valores: privacidad y seguridad; derechos humanos; no maleficencia; confiabilidad; exactitud y precisión; relevancia; eficiencia y usabilidad, y transparencia.		
Procesamiento			
Almacenamiento			
Análisis			
Compartición			
Archivo/eliminación			

Fuente: elaboración propia.

Cuando se reflexiona sobre la manera como ha trabajado el SETE en sus dos años de funcionamiento respecto a las operaciones estadísticas, se evidencia que recoge de forma implícita los dos modelos mencionado. En relación con el modelo de Davis, en el SETE se considera que cada dos años se haga una evaluación de los ejes y los principios que configuran el marco ético, lo cual corresponde con la fase de investigación de Davis. La fase de análisis corresponde a todo el proceso de evaluación de la operación estadística. La fase de articulación es análoga al plan de acción concertado que se hace entre el grupo de trabajo y los encargados de la operación estadística. La fase de la acción se corresponde con las acciones realizadas por los encargados por la operación y que tienen un seguimiento por parte del grupo de trabajo del SETE<sup>2</sup>. Por otra parte, las fases del ciclo de vida del modelo del BID están ya incluidos dentro

del modelo GAMS0 -GSBPM y que contiene las siguientes etapas: identificación de necesidades, diseño, construcción, recolección, procesamiento, análisis, difusión y evaluación.

### 1.4. Aplicación del modelo para usos no estadísticos

Este marco presenta guías generales para una gestión ética de los datos, así como consideraciones específicas que solo son aplicables al momento de tratar datos estadísticos. En general, las obligaciones para los diferentes actores del SEN presentes en la Ley 2335 por la cual se expiden disposiciones sobre las estadísticas oficiales en el país (Congreso de la República de Colombia, 2023) deben considerarse relevantes exclusivamente para este tipo de uso. Sin

<sup>2</sup> Para conocer más sobre el funcionamiento del SETE, sus procesos de evaluación y el documento maestro en el que se plantea el Marco ético y los órganos que lo conforman, consultar: <http://www.sen.gov.co/conozca-el-sen/instancias/sete>

embargo, este documento tiene como fuentes referentes normativos, académicos y de trabajo relevantes para todos los campos del trabajo de los datos. Entre ellos se resaltan el *Enfoque para la programación basado en los derechos humanos* de las Naciones Unidas (2022), el *Reglamento General de Protección de Datos* y la *Data Act* de la Unión Europea (2016 y 2022), el *Plan Nacional de Infraestructura de Datos* de MinTIC, DNP y Presidencia de la República de Colombia (2021) y el *Marco Ético para la Inteligencia Artificial en Colombia* del Gobierno de Colombia (2021) y todos sus planteamientos han sido construidos a través del análisis e integración de lo allí planteado.

Por ello, para usos comerciales, investigativos, informativos, archivísticos o artísticos, entre otros (de aquí en adelante llamados “uso general”), se deberán tener en cuenta los presentes principios para contar con una perspectiva ética en todo su ciclo de vida. Sin embargo, dos características del uso estadístico lo hacen diferente del uso general: la obligatoriedad en la entrega de la Información (Ley 2335, artículos 19, 22, 30) y la reserva estadística (Ley 2335, Cap. V). También cabe aclarar que hay otros dos usos de los datos que, pueden tener consideraciones particulares, por lo que requieren una regulación específica y no pueden agruparse con el uso general: el uso judicial y el uso para inteligencia, defensa y seguridad nacional.

Para el uso general, al no existir una obligación por parte de los usuarios para compartirlos u otorgarlos aplicará la Ley 1266 de 2008 por la que se regula el manejo de la información contenidas en bases de datos personales, exceptuando los límites en ella misma impuesta. De este modo, asuntos como el consentimiento para la generación y el uso de los datos, la imposibilidad de extender el uso de los datos más allá de las autorizaciones presentadas, el derecho de los individuos para que sus datos sean eliminados de estudios o procesos, la propiedad de los datos y los límites para compartir, transmitir y comercializar información deberán ser tenidos en cuenta.

Por su parte, el uso general no cuenta con la figura de la reserva estadística que exige que las publicaciones impidan la identificación de sujetos individuales y la transmisión de la información con otros fines que el estadístico, y conlleva un compromiso de confidencialidad y seguridad por parte de todos los actores involucrados. Por ello, en los usos generales se deberán constituir mecanismos propios que permitan mitigar los riesgos para los principios de privacidad y seguridad, los derechos humanos y no maleficencia que pueden presentarse ante la ausencia de esta figura.

Pese a lo anterior, se puede evidenciar que el uso general de los datos implica fortalecer y adaptar los estándares de los principios que se presentarán a continuación. Una gestión ética de los datos requiere dar cuenta de cada uno de estos principios respondiendo a las necesidades dadas por el contexto de su uso específico. El uso estadístico es una particularización en la que se siguen los principios generales de este marco.

# 2

## Principios éticos de la gestión ética de los datos



Exhibit 3: US economic growth expected to peak this quarter as of May 5, 2021



Para la elaboración de los principios que aquí se proponen, se tomaron como referencia las siguientes iniciativas: el Reglamento general de datos de la Unión Europea RGPD (2016), la Data Act de la Unión Europea (2022), el Marco ético de la Inteligencia Artificial para Colombia (2021), la Gestión ética de los datos del Banco Interamericano de Desarrollo (2019), el Core values of official statistics de UNECE (2022), Comisión Económica de la Unión Europea (2021), la Guía de estándares de interoperabilidad Gobierno en línea Colombia (S.F) y la Ley 2335 (2023). En estas iniciativas se encontraron en total 44 principios. Después de analizar cada una de las definiciones de estos principios se agruparon en siete grandes bloques, de acuerdo con su grado de afinidad por lo que cada bloque representa un principio.

A continuación, se le dio el nombre que se consideró más adecuado para cada principio, y se procedió a hacer su respectiva definición. Para ello, se tuvo en cuenta lo que señalaban las iniciativas mencionadas y la literatura adicional, de carácter conceptual, sobre el tema. Cada una de estas definiciones fue discutida por el grupo de trabajo en dos rondas, de tal manera que en la primera se planteó la definición de los principios acordados y, en la segunda, se discutieron los ajustes necesarios para garantizar que dichas definiciones atendiesen todas las consideraciones presentes en los referentes. A continuación, se presentan dichos principios con sus definiciones.

1. Privacidad y seguridad.
2. Derechos humanos.
3. No maleficencia.
4. Confiabilidad: precisión y exactitud.
5. Relevancia.
6. Eficiencia y usabilidad.
7. Transparencia.

### 2.1. Privacidad y seguridad

Se ha vuelto un lugar común señalar que uno de los efectos negativos que ha tenido la revolución del big data ha sido la erosión de la privacidad de las personas. Como bien lo señalan Gry Hasselbalc y Pernille Tranberg (2016), “las repercusiones sociales del manejo de datos han sido enormes sobre el funcionamiento de las democracias, pero, sobre todo, en la esfera privada de las personas” (2016, pos. 62). Estos mismos autores tienen una definición de la privacidad que es muy útil para entender el alcance que tiene este valor sobre el manejo ético de los datos. A este respecto afirman: “la privacidad es el control sobre los propios datos y el derecho a decidir quién sabe que acerca de usted y cuando” (2016, pos. 2533).

En el marco de una democracia liberal, se considera como un elemento esencial la separación entre la esfera de lo público y la esfera de lo privado. La primera corresponde a la vida en común, a la construcción de un proyecto colectivo y a la solución de problemas que solamente se pueden resolver con la cooperación de los ciudadanos que hacen parte de una comunidad. Lo privado, en cambio, refiere al proyecto de vida de cada uno de nosotros e incluye relaciones estrechas con los que se consideran más cercanos. En este tipo de sociedad, uno de los derechos fundamentales es el derecho a la intimidad, que refiere al derecho de cada persona de decidir sobre qué vida quiere vivir y con quiénes quiere tener una relación estrecha. Como todos los derechos, el derecho a la intimidad tiene un límite, que es justamente el de no hacer daño a los otros. Ahora bien, es parte central de este derecho la potestad de dar información sobre los asuntos de carácter íntimo a otras personas y a la sociedad en general; es ello a lo que refiere la expresión “control sobre los propios datos” mencionada en la definición citada. De acuerdo con esto, la privacidad no significa que hay una prohibición absoluta de recolectar información sobre la esfera íntima de las personas; estos se pueden obtener siempre y cuando se cuente con la aquiescencia de la persona de quien se requiere la información.

Cabe resaltar que en todos los procesos de recolección estadística se parte de un Principio de Confidencialidad sobre la información recolectada (Ley 2335, art. 4, numeral 12 y 13), por lo que en la adquisición de datos privados o personales con fines estadísticos se establece un compromiso con la fuente de información para el respeto y la garantía de su privacidad. Este compromiso se ve representado en la Reserva Estadística que es “la obligación legal del DANE y de las entidades del SEN, en el marco de la producción estadística, de garantizar que los datos que impliquen la identificación directa o por deducción de las fuentes primarias o secundarias de personas naturales o jurídicas estén restringidos al público en general” (Ley 2335, art. 5, numeral 18). De este modo, aun cuando las entidades adscritas al SEN, así como todos los ciudadanos, se encuentran en la obligación de compartir su información ante la autoridad estadística, esta se encontrará protegida, salvaguardada y es una obligación central presentarla de tal modo que no se pueda asociar con ningún tipo de individuo.

Ahora bien, como lo muestra de manera contundente Carissa Véliz, (2021), las grandes empresas que tienen el dominio sobre las redes sociales han vuelto el acceso a los datos privados de las personas un negocio. Ello implica un grave peligro para la libertad de las personas, que al ser seres sintientes, vulnerables y dependientes pueden ser manipulados y presionados por los otros, lo cual puede implicar el hacer cosas que van en contravía de sus mejores intereses. Por ello, es fundamental para la privacidad que las personas provean la información de manera voluntaria y que el uso que se haga de la información no atente contra sus intereses. Además, un apoyo fundamental a la privacidad es que se brinde a las personas información veraz y oportuna, de tal manera que tengan más control sobre sí mismas y sobre las decisiones que tomen.

Siguiendo a Mike Loukides y Hilary Mason (2018), para garantizar la privacidad se deben tener en cuenta los siguientes cinco lineamientos:

1. El consentimiento de las personas en torno a la información que brindan y a cómo esta se utiliza, reconociendo los derechos que le asisten y los mecanismos existentes para ejercerlos. En particular hacer explícito el cumplimiento del Principio de confidencialidad y de Reserva Estadística en este tipo de operaciones.
2. La claridad en relación con la información que las personas dan y reciben.
3. Consistencia y confianza en el manejo de los datos.
4. El control de las personas sobre los datos que proveen.
5. El conocimiento de las personas sobre el riesgo que puede haber sobre ellas por la información que tanto dan como reciben.

En relación con los efectos sobre la seguridad, deben considerarse dos asuntos que deben ser considerados por quienes lleven a cabo labores estadísticas y de recolección de datos: la seguridad de los datos y la seguridad de las personas y las poblaciones.

La búsqueda de la seguridad de los datos es un asunto técnico y de procedimientos que involucra directamente a todos los actores involucrados. Además, es un proceso donde el eslabón más débil, en muchas ocasiones, es el ser humano. En el contexto contemporáneo, y reconociendo las necesidades expresadas en el principio de privacidad, se entiende la seguridad como el conjunto de medidas que pueden tomarse con el fin de garantizar que los datos solo puedan ser accedidos o consultados, para su conocimiento, uso o manipulación, por las personas debidamente autorizadas en los contextos para los que han sido recolectados. Así, se deberán tomar medidas que prevengan la fuga o la pérdida de los datos ya sea por ataques informáticos o por acción de las personas que tienen acceso a los mismos. Igualmente, se debe tener en cuenta que

hay intereses económicos que convierten a los datos en un blanco para este tipo de ataques, ya sea con el fin de extender las bases de datos con fines comerciales, para llevar a cabo acciones de suplantación de identidad o para secuestrar los datos demandando una retribución económica por su liberación. Por ello, se deberá contar con la asesoría de especialistas en el tema que permitan seguir las mejores prácticas en seguridad informática.

Un riesgo adicional que traen los contextos informáticos para los datos es garantizar su adecuado almacenamiento. El final del ciclo de vida de los datos, descrito en la sección 1.2. y el anexo A, concluye con el archivo o la eliminación de los datos adquiridos. Cuando el final sea la eliminación se deberá garantizar que no existan copias adicionales ni la posibilidad de acceder a estos datos por ningún medio informático. En el caso del archivo se deberá garantizar que se mitiguen los riesgos por fallas o deterioro de hardware y que los datos se encontrarán guardados en formatos que permitan su acceso a futuro.

Finalmente, pero no menos importante, la seguridad en las operaciones de datos también implica una protección de los derechos humanos. Al diseñar y llevar a cabo acciones estadísticas se deberá reconocer la existencia de poblaciones vulnerables para quienes la divulgación de la información puede representar un riesgo para sus vidas. Datos sensibles como la geolocalización de los individuos, la pertenencia a comunidades y colectivos, la participación en actividades particulares, entre otros, podrán poner en riesgo a los sujetos dependiendo del contexto donde el dato surja y sea usado. Estas consideraciones para la seguridad de los participantes deberán ser tenidas en cuenta antes de la realización de cualquier estudio y se deberán implementar los cuidados necesarios.

Algunas de las recomendaciones técnicas que presenta el BID son:

- “Considerar medidas, tales como políticas de respaldo (hacer una o varias copias que permitan recuperar la información en caso de pérdida) de la información, medidas de

encriptación, accesos diferenciados según el perfil, y protección contra la piratería informática” (p. 44).

- Los “cortafuegos, el cifrado, la anonimización y la codificación son algunas de las modalidades de protección de los datos personales” (p. 14).

- “A menudo las fallas de seguridad son producto de la acción humana”.

- “Los ficheros deben ser protegidos adecuadamente contra riesgos naturales y humanos”.

- La protección implica “seguridad de los datos, privacidad de los usuarios, anonimato y acceso por parte de aquellos a quienes hace referencia” (p. 26).

En líneas generales, las buenas prácticas de seguridad informática, así como de custodia y acceso a los datos, son referidas como un componente de la privacidad (la garantía de que solo los indicados puedan acceder a los datos) o como un componente independiente (como una propiedad intrínseca de los sistemas y mecanismos informáticos). Todo medio tecnológico utilizado en el manejo de los datos cumplirá con las principales recomendaciones y buenas prácticas en seguridad informática, esto con el fin de evitar usos no autorizados ni divulgación no intencional de los datos.

## 2.2. Derechos humanos

Los derechos humanos son el referente normativo tanto para los sistemas jurídicos nacionales, como para las relaciones entre países. Además de ello, los derechos humanos son derechos morales y se han convertido en el eje central del ethos de las sociedades democráticas. Estos derechos protegen intereses morales legítimos de los ciudadanos que se justifican en valores como la libertad, la equidad y la solidaridad.

En la medida en que los derechos humanos son derechos morales es importante tener en cuenta los siguientes aspectos: en pri-



mer lugar, como lo señala Kant y, en el siglo pasado, Hannah Arendt, las personas, sin importar su condición, tienen el derecho a tener derechos. Esto quiere decir que tiene que haber, dentro del sistema jurídico, mecanismos efectivos a los que las personas puedan acudir para hacer valer sus derechos. En segundo lugar, para las personas que tienen capacidad de tomar decisiones, la protección de sus derechos va acompañada de su obligación de respetar los derechos de los otros. En otras palabras, a la base de los derechos está la reciprocidad y el reconocimiento mutuo entre los miembros de la comunidad política. En tercer lugar, si bien garantizar el cumplimiento de los derechos requiere de un Estado y de un sistema jurídico, en un estado democrático los ciudadanos asumen, como parte de su deber moral, velar porque estos se respeten.

Bajo este marco, el uso estadístico de los datos tiene como uno de sus propósitos promover la defensa de los derechos humanos y garantizar los derechos de las personas en la manera en que se producen y se difunden los datos que se les requieren y proveen.

Ahora, los derechos humanos no deben concebirse solamente como un principio moral, sino también como un enfoque conceptual que las incitativas de desarrollo sostenible deben tener en cuenta. El propósito de este enfoque es que a través del análisis de las desigualdades se pueden corregir las prácticas discriminatorias y las distribuciones injustas de poder que impiden el progreso y el desarrollo de todos.

Este enfoque comprende un sistema de derechos y obligaciones establecidos por el derecho internacional, en el que se incluyen todos los derechos civiles, culturales, económicos, políticos y sociales, y el derecho al desarrollo. Además, este enfoque se soporta sobre algunos de los principios como: la universalidad, la indivisibilidad, la igualdad, la no discriminación, la participación y la rendición de cuentas.

La Organización de las Naciones Unidas (ONU) ha sintetizado los esfuerzos para promover los derechos humanos, en el imperativo moral de no dejar a nadie atrás, lo

que implica combatir las desigualdades y la discriminación a través del esfuerzo mancomunado de las naciones. Este principio es el que se encuentra en el corazón de la Agenda 2030 para el Desarrollo Sostenible.

La prevención de la estigmatización y la discriminación, relacionada con la autonomía y la dignidad humana es una exigencia ética que tiene una relación directa con la producción estadística. Esto se ve reflejado en el hecho de que, la imparcialidad y la inclusión son principios de las estadísticas oficiales (Ley 2335, art. 4, numeral 3 y 4) y la dignidad, la confidencialidad y la transparencia son un eje de articulación del *Marco Ético para las Operaciones Estadísticas del DANE* (DANE, 2020). Así pues, en los usos de los datos con fines estadísticos, el respeto a los derechos humanos es un principio que no se puede soslayar y es una de sus razones de ser. Sin embargo, para lograr este propósito es menester que en todas en todas las fases del ciclo de vida de los datos se garantice al máximo la protección de los derechos fundamentales de todas las personas naturales y jurídicas.

### 2.3. No maleficencia

Como bien señalan Tom Beauchamp y Walters Leroy, hay un principio ético muy antiguo de la práctica médica que ha sido reivindicado por el utilitarismo en todas sus vertientes que dice “*primum non nocere*”, “*primero no hacer daño*” (Beauchamp, Tom, Walters, Leroy, 1999, p. 20). El principio señala que cuando una persona decide algo debe hacerlo de tal manera que las consecuencias directas e indirectas de su acción no generen daños en otros seres que merecen respeto y consideración. En los últimos años, en la sociedad se ha aceptado que quienes merecen respeto y consideración no son solamente los seres humanos, sino también los animales y la naturaleza en general. Este principio está a la base del concepto de riesgo, entendido este como la probabilidad de que el efecto de una acción genere un daño sobre algo o alguien. Ahora bien, en virtud de que los seres humanos forman una red amplia y densa de relaciones entre ellos y con la naturaleza, no es factible lograr en

todo tiempo y en todo lugar acciones que no puedan generar algún daño, es decir, que no tengan un riesgo. Por ello, el principio del no daño debe entenderse en un sentido menos radical que consistiría en la exigencia de realizar acciones que tengan el menor daño posible. Un ejemplo claro de los daños que la generación de datos e información en el contexto de la producción estadística es la carga excesiva de las fuentes y, por ello, se considera indica como para cumplir con los principios estadísticas oficiales las “entidades productoras de estadísticas controlarán la carga que supone responder y fijarán objetivos para reducirla progresivamente” (Ley 2335, art. 4, numeral 11).

Por otra parte, ligado a este principio del no daño, se deriva otra exigencia y es que en caso de que la acción produzca un daño sobre otros, debe tratarse de compensar a la persona o grupo afectado y tratar de mitigar razonablemente las pérdidas ocasionadas. Esta exigencia implica de parte de las organizaciones que tengan mecanismos para gestionar los riesgos de tal manera que haya formas de prevenirlos y en caso de que se presenten se deben mitigar y compensarlos.

Estos dos principios, el de minimizar el riesgo y el de compensarlo, son fundamentales en el manejo de los datos. A este respecto, Mike Loukides, Hilary Mason y D.J. Patil, sugieren que en toda organización el manejo de los datos tiene que atender seriamente a la cuestión de “si los datos recogidos podrían causar daño a un individuo o grupo” (Loukides, Mike, Mason, Hilary, y Patil, D.J. *Ethics and data science*, 2018, p. 24). Identificar los posibles daños que se puedan dar en la recolección y el manejo de los datos es el centro de lo que hoy se denomina la gestión de los riesgos de la información. El propósito de este tipo de gestión, desde un punto de vista ético, es defender el buen nombre de la organización y velar al máximo porque la información que la organiza-

a las personas, los grupos y a todo aquello que tenga un valor moral, como son la naturaleza y la cultura.

### 2.4. Confiabilidad: exactitud y precisión

Un atributo ético fundamental de la información estadística es la confiabilidad. En concordancia con lo propuesto en el *Plan Nacional de Infraestructura de Datos (PNIE)*<sup>3</sup>, que en su cuarto principio (Confianza pública y gestión ética de los datos) recomienda la integración de la gestión ética de datos, se propone que uno de los pilares del SEN debe ser el robustecimiento de la confiabilidad que, a su vez, da lugar a la confianza razonable de todos los actores involucrados y la sociedad civil en general.

Siguiendo los postulados de Onora O’Neill, respecto de la confiabilidad (2018; 2017), se concluye que la confiabilidad, además de la honestidad de la persona o la entidad (relacionada con la transparencia), precisa de la demostración de competencia en el sentido de la capacidad de llevar a cabo una tarea de forma certera. Así como de fiabilidad para que se genera una alta probabilidad de que se entreguen buenos resultados de forma consistente. En este sentido, se puede afirmar que la confiabilidad de cualquier forma de medición depende de su exactitud, precisión y rigurosidad técnica, principios fundamentales de las estadísticas oficiales (Ley 2335, art. 4, numeral 2 y 8).

Por una parte, la exactitud se entiende como la cercanía o la coincidencia entre el valor real y el valor medido. En otras palabras, hablar de exactitud hace referencia a qué tanto se acerca una medición o un juicio a la realidad del objeto observado.

Según la *Guía de estándares de calidad e interoperabilidad de los datos abiertos del gobierno de Colombia*<sup>4</sup>, la exactitud se re-

<sup>3</sup> [https://www.MinTIC.gov.co/porta1/715/articles-179710\\_recurso\\_2.pdf](https://www.MinTIC.gov.co/porta1/715/articles-179710_recurso_2.pdf)

<sup>4</sup> Disponible en Microsoft Word - Documento PNID Versión para comentarios.docx (MinTIC.gov.co)

fiere a “datos diligenciados correctamente”. La guía propone que “los datos abiertos a los usuarios estén diligenciados correctamente y que a su vez brinden información exacta en sus conjuntos de datos” (p. 27). Adicionalmente, la Norma ISO/IEC 25012 de 2008 define la exactitud como “el grado en el que los datos representan correctamente el verdadero valor del atributo deseado de un concepto o evento en un contexto de uso específico” (ISO 25012 (iso25000.com)). De esta manera, propender por la exactitud en los datos es la búsqueda de que representen correctamente su objeto de estudio.

Por otra parte, la precisión se entiende como el grado en el que un instrumento de medida arroja valores iguales o similares en cada medición. Así, un instrumento de medición preciso, independientemente de su exactitud, arrojará resultados consistentes. La formación de una cultura de la confiabilidad y de la confianza que debe seguirle, necesita de la exactitud empírica de los datos obtenidos, la demostración de profesionalismo y del compromiso con la materialización constante de datos exactos. En otras palabras, para la materialización de la confiabilidad se necesita que todos los actores del SEN muestren, en su actuar, que son predecibles en la entrega de información evitando, tanto como sea posible, la dispersión de resultados.

Igualmente, al pensar en la forma en que la exactitud y la precisión promueven la confiabilidad de las estadísticas, se encuentra que teniendo mediciones que se correspondan de manera fidedigna con la realidad que describe se logra satisfacer la necesidad de competencia, entendida como la capacidad de llevar a cabo mediciones de calidad. Asimismo, la precisión genera fiabilidad dado buenas probabilidades de que las ventajas de la exactitud se brindarán de forma consistente o, en otros términos, fiable. Una manera para lograr la confiabilidad en las estadísticas oficiales es promover la “coherencia y comparabilidad [esto es, que] las estadísticas deberán ser acordes con los estándares internacionales, de manera que sean comparables a lo largo del tiempo y con los demás países, y coherentes en su

marco conceptual, metodologías aplicadas y resultados producidos” (Ley 2335, art. 4, numeral 1) ya que su aplicación, tanto para la producción de los datos como la construcción de información, conserva la perspectiva de estudio a lo largo del ejercicio estadístico. En consecuencia, el principio de confiabilidad exige que, en el marco del SEN, haya información exacta que se produzca de manera consistente.

### 2.5. Relevancia

La relevancia debe entenderse como un concepto relacional. Es así como la información nunca es relevante en sentido absoluto sino solo en relación con un problema puntual de importancia para una sociedad. En particular, se debe entender la relevancia como el opuesto de la casualidad o el azar en los que, por definición, no se pueden prever, no hay una relación que ate o junte las variables y los elementos involucrados de manera significativa. Por el contrario, la información es relevante para un problema en tanto que permite un mejor entendimiento de la situación sujeta de observación, al aumentar la capacidad explicativa o predictiva de dichas situaciones. De esta manera, la relevancia constituye un pilar fundamental de la generación de valor para los actores involucrados y para la población en general.

En gran medida, el valor de cualquier medición o estudio depende de la relevancia, la pertinencia y la oportunidad de la información y las relaciones que postula. En este sentido, el cuidado de la relevancia minimiza los falsos positivos, las correlaciones débiles y la información que por su contenido, su forma y su tiempo de difusión no contribuye al mejoramiento de la comprensión y a la toma de decisiones relacionadas con esta, a través, de la justificación y la explicación de las preguntas, los métodos y los resultados.

La definición de relevancia presentada es afín con la *Guía de estándares de calidad e interoperabilidad de los datos abiertos del gobierno de Colombia*, que postula que la relevancia debe entenderse en términos de que “los datos publicados deben ser de utili-

dad para los usuarios”(2020, p. 8), por lo que el criterio de relevancia implica el trabajo de determinar cuidadosamente “los temas importantes, destacados, significativos o con mayor demanda por parte de los usuarios que se deben publicar en la apertura de datos” (p. 17).

En resumen, y como se postula en la *Norma Técnica de la Calidad del Proceso Estadístico (2020)*<sup>5</sup>, la relevancia “[s]e refiere al grado en que las estadísticas satisfacen las necesidades de información de los usuarios” (p.13). Al considerar que el SEN se propone, en el marco de sus objetivos, el suministro de estadísticas oficiales y la promoción de su conocimiento y uso, se hace evidente que la producción y el aprovechamiento de los datos y la información disponible dependen del aseguramiento de la relevancia para sus usuarios potenciales. Por ello, los resultados de los procesos estadísticos deberán ser difundidos de tal forma que puedan ser conocidos y aprovechados por los todos actores interesados.

### 2.6. Eficiencia o buen uso

El aprovechamiento de fuentes secundarias para la producción de información y la capacidad de generar valor público a través de dicha información depende en gran medida de la oportunidad, la exactitud y la relevancia de los datos, atributos que presentan una relación directa con la facilidad y la transparencia en el acceso a estos. La eficiencia se puede entender como la definición precisa de mecanismos y canales que permitan realizar una adecuada gestión de las solicitudes de acceso a los datos y que propicien una compartición oportuna, segura, sistemática y objetiva de los mismos al interior de las entidades, con entidades externas del sector

público o privado o con la academia, evitando duplicidades y promoviendo economías de escala. Los requisitos para el reúso de los datos deben entenderse en dos vías: por una parte, el aprovechamiento de estas fuentes secundarias cuando tengan lugar y, por otra, las consideraciones sobre la propia producción para que esta pueda ser utilizada por los demás actores.

La eficiencia como eje central en el intercambio y la interoperabilidad de los datos requiere etapas clave asociadas a la anonimización, la privacidad diferencial, la transparencia y la objetividad en el trámite de las solicitudes de acceso, de tal forma que se dé cabal cumplimiento a los demás principios mencionados anteriormente. En este orden de ideas, la eficiencia se puede configurar como un estándar por defecto que parte de un mapeo, tanto de los datos como de los potenciales usuarios, y que dé cumplimiento a lo establecido en el artículo 333 de la Ley 1955 de 2019, el Decreto 2106 de y el Decreto 2052 de 2020<sup>6</sup>.

En la actualidad, múltiples sistemas, dispositivos e instituciones se encuentran en constante producción de datos derivados de la actividad de las personas. Para llevar a cabo su registro, quienes lo hacen deben contar con una autorización para el uso particular para el que se proponen. Asimismo, y con el fin de garantizar que estos datos puedan ser utilizados a cabalidad y no se caiga en procesos redundantes en las instituciones, estos deberán ser almacenados de tal manera que puedan alimentar múltiples procesos relevantes y autorizados. En este caso es clave recordar que existen usos considerados legítimos en la normatividad sin importar el origen inicial de los datos, tales como los fines de archivo en interés público, la seguridad nacional, los fines de

<sup>5</sup> Disponible en NORMA TÉCNICA NTC COLOMBIANA 947 -1 (dane.gov.co).

<sup>6</sup> Artículo 333 de la Ley 1955 de 2019 “Supresión de trámites, procesos y procedimientos para la modernización y eficiencia de la administración pública”, el Decreto 2106 de 2019 “Por el cual se dictan normas para simplificar, suprimir y reformar trámites, procesos y procedimientos innecesarios existentes en la administración pública” y el Decreto 2052 de 2020 “Por medio de la cual se establecen disposiciones transversales a la rama ejecutiva del nivel nacional y territorial y a los particulares que cumplan funciones públicas y/o administrativas, en relación con la racionalización de trámites y se dictan otras disposiciones”.

investigación científica e histórica o los fines estadísticos.

Las condiciones para el buen uso de los datos se componen de dos elementos: las condiciones técnicas que tienen los datos para poder ser usados por los interesados y los límites que se tienen para el uso de los datos de acuerdo con su origen.

Para cada uno de estos dos niveles se deben tener en cuenta los siguientes estándares:

- A nivel técnico, se debe observar la *Guía de estándares de calidad e interoperabilidad de los datos abiertos del gobierno de Colombia*.

- En particular, se deberá recordar que los datos recolectados existen para más que un asunto o estudio particular y deberán ser almacenados de tal modo que puedan comprendidos e interpretados en otros usos legítimos.

- Aquellas entidades interesadas y autorizadas para llevar a cabo los usos legítimos contemplados en cada dato deberán colaborar entre ellas.

- Si bien los datos pueden ser compartidos o divulgados de manera abierta al interior de una misma institución o con entidades externas, antes de hacerlo es importante tomar las medidas necesarias para garantizar que solo puedan ser usados para los fines establecidos en su recolección.

- El manejo ético de los datos, tal y como se contempla en el presente documento, debe darse durante todo su ciclo de vida: desde la recolección hasta su archivo o eliminación, pasando por las etapas de almacenamiento, análisis y compartición (Buenadicha, et. al, 2019).

- Cuando los datos sean usados para la toma de decisión por parte de entidades o instituciones, el criterio y el proceso de toma de decisión deberá ser comprensible, ajustable, analizable y, finalmente, reversible por parte un humano involucrado en el proceso. Si bien se comprende el valor de los sistemas de inteligencia artificial como apoyo en la toma de decisión, estas no podrán ser dadas desde una caja negra<sup>7</sup>.

---

<sup>7</sup> Se considera que un sistema opera como una caja negra cuando existe opacidad o falta de claridad sobre los criterios y los pesos para la toma de decisión por parte de un algoritmo. Si solamente se reconocen las entradas y las salidas del sistema y se presupone que los procesos son correctos, pero se desconoce, en este caso se está operando desde una caja negra.

### 2.7. Transparencia

La transparencia se puede entender como la cualidad o la virtud deseable en el ciclo de vida de los datos que permite conocer de forma fiable los procedimientos, las técnicas y los usos que se presentarán durante el ciclo. La transparencia también concierne a toda “la información sobre las fuentes, los métodos y procedimientos aplicados en la actividad estadística” (Ley 2335, art. 4, numeral 9) que deben ser de conocimiento público. Cabe aclarar que ser transparente respecto a las diferentes acciones llevadas a cabo durante el ciclo de vida o la actividad estadística no debe minar de ninguna manera la protección de los datos, por lo que se deberá seguir haciendo énfasis en el respeto del primer principio, Privacidad y Seguridad.

Un aspecto clave de la transparencia es comunicar, de manera clara y explícita, el alcance y las limitaciones de los conjuntos de datos. De esta manera, no solo se contribuye a la confianza en los datos y los resultados de los procesos, también permite la correcta reutilización y el uso compartido de los datos. Por ello, la transparencia implica una contextualización adecuada que permita a cualquier actor que no se encuentre involucrado directamente en la producción de los datos el comprender lo que ellos pueden lograr.

Asimismo, en la medida en que se hace uso primario y secundario de los datos con fines estadísticos, comunicar e informar los procesos detrás de los datos (acceso, diseño, procesos analíticos y tecnológicos empleados y su difusión) permite su aprovechamiento integral, lo que genera credibilidad y confianza con la plena garantía de contar con datos limpios (libres de manipulación e intereses particulares) y que velan en proteger aquella información que pueda afectar la garantía a la intimidad e integridad de las personas. La transparencia como ejercicio de responsabilidad social garantiza también el acceso equitativo a la información pública derivada del aprovechamiento de un conjunto de datos con fines estadísticos.

# 3

# Diseño institucional de la evaluación de la gestión ética de los datos del SEN



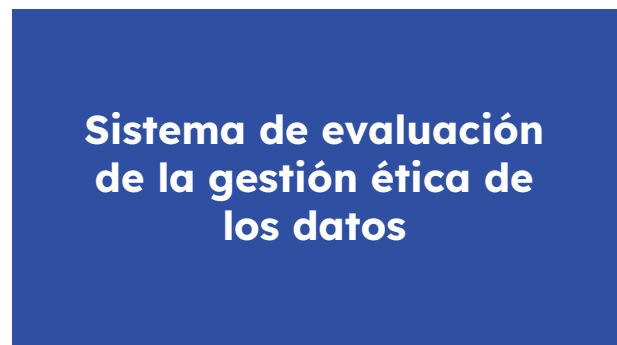
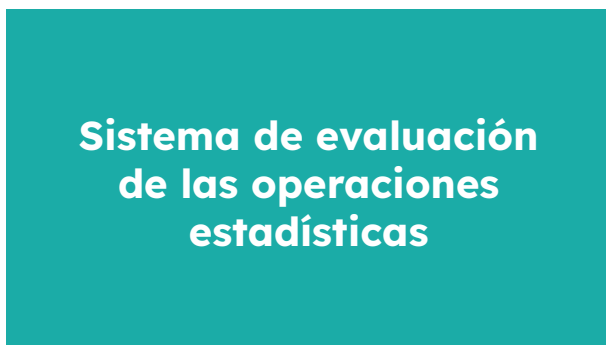
## Marco ético de los datos del Sistema Estadístico Nacional

Con el fin de garantizar que los datos, que, con fines estadísticos, generen y difundan las entidades que hacen parte del SEN, cumplan con los estándares éticos es fundamental contar con un arreglo institucional que permita hacer seguimiento por parte del CAD que está justamente encargado de velar por la calidad de dichos datos en el SEN. De esta manera el presente marco facilita el cumplimiento de las obligaciones de los miembros del SEN (Ley 2335, art. 13) y se permite el alcance de una de las finalidades del SEN: “propiciar intercambios de información entre quienes lo conforman para una producción eficiente y a fomentar la cultura estadística, de manera que a tra-

vés de él se contribuya a la apropiación de la información estadística en la sociedad, garantizando el uso ético y adecuado de los datos individuales que sean gestionados en el sistema” (Ley 2335, art 10).

Es importante aclarar aquí que habría dos sistemas de evaluación interrelacionados: el primero de ellos corresponde al actual SETE, que está encargado de la evaluación de las operaciones estadísticas, y que actualmente está en funcionamiento; el segundo es el Sistema de evaluación de la gestión ética de los datos. Este documento se centra en el segundo sistema.

**Gráfica 2.** Sistemas de evaluación



Fuente: elaboración propia.

Dada la multiplicidad de entidades que hacen parte del SEN, la tarea de seguimiento no es fácil. No basta que los actores del sistema se comprometan con tener sistemas de gestión que cumplan con los principios como la privacidad, la seguridad, la transparencia, etc. Como lo han mostrado diversos estudios sobre la ética en las organizaciones, la existencia de códigos éticos o las declaraciones de principios no garantizan que las prácticas organizacionales y específicamente las que tiene que ver con la gestión de datos respondan a los estándares éticos allí establecidos (García-Marzá, 2004). Frente a ello, se podría pensar en una forma de seguimiento en el que, desde una instancia articuladora o reguladora como

sería el CAD, se hiciera una evaluación ética de cómo se han producido los datos que cada entidad provee al SEN. El problema de esta alternativa es que requeriría de una instancia centralizada encargada de la evaluación ética de la producción de información de cada entidad. Por ejemplo, frente a los datos que la DIAN provea información sobre el recaudo de cierto impuesto se tendría que hacer una evaluación ética sobre cómo se recolectó y proceso esta información. Como se dijo antes, dado el número de entidades que hacen parte del sistema y teniendo la diversidad de datos que cada entidad provee al SEN, este proceso se puede volver engorroso y al ser completamente centralizado puede llevar a una interferen-



cia que puede generar ineficiencias en la gestión de los datos en las entidades.

En lugar de ello, lo que se propone aquí es una estrategia ligada a una forma de gobernanza descentralizada en la que, en lugar de evaluar cómo se produce cada dato, se evalúa si los sistemas de gestión de la información de las entidades cuentan, por una parte, con un marco compatible con los principios éticos aquí establecidos y, por otra, con mecanismos de seguimiento para velar que dichos principios se cumplan al interior de la entidad. Para evitar el centralismo, se seguiría un mecanismo similar al que usan iniciativas como el Global Reporting Initiative (GRI)<sup>8</sup>, en el que las organizaciones hacen un proceso de autoevaluación respaldada con evidencias que después es presentada a una instancia que, partiendo de una verificación de estas evidencias, hace una evaluación de la organización y luego genera un aval en el que se señala que la entidad cuenta con procesos acordes con los principios del GRI. Además, se propone la existencia de instancias que verifiquen, evalúen y avalen el cumplimiento de los principios del marco ético por parte de las entidades involucradas.

Para entender mejor en qué consiste la propuesta a continuación se presentan los lineamientos para seleccionar las entidades a evaluar, los componentes del sistema y el proceso.

### 3.1. Lineamientos para la selección

Dado que el número de entidades que hacen parte del SEN es bastante grande (aproximadamente 113 entidades) no es factible hacer una evaluación de todas ellas en un tiempo razonable. Por tal motivo, es necesario considerar los siguientes criterios de selección:

**1. Se tendrían en cuenta las necesidades de información que el CAD, en sus mesas de trabajo, considere estratégicas y que requieran de intercambio de datos entre entidades.** En este sentido, se escogerían las entidades que puedan proveer los datos que atiendan de manera más apropiada a dichas necesidades de información.

**2. La madurez que tengan las entidades en la gestión de datos.** Es mejor evaluar aquellas entidades que tengan procesos de gestión de datos ya consolidados, es decir, que se tengan procesos claros y consistentes para el ciclo de vida de los datos, que de acuerdo con el Plan Nacional de Infraestructura de Datos; este contiene las siguientes etapas: creación, procesamiento, análisis, almacenamiento, intercambio, uso y análisis de datos, archivo y preservación y reutilización (MinTIC, Plan de Infraestructura Nacional de Datos, Anexo, Ciclo de vida de los datos, 2022).

**3. La voluntariedad de las entidades.** Solo se evaluarían las entidades que quieran someterse al proceso.

### 3.2. Componentes

Antes de presentar los componentes del sistema, es necesario mencionar los principios que guían su composición y funcionamiento. Dichos principios son la independencia, la transparencia, la interdisciplinariedad y la agilidad.

**Independencia:** dado que cada uno de los componentes del sistema está en función de velar porque los principios del marco ético se cumplan es fundamental que las evaluaciones que se hagan estén libres de influencia por parte de intereses personales o de reputación institucional. Lo central es que los datos que se provean al sistema sean verídicos, oportunos, precisos, confiables y garanticen al máximo la privacidad y los derechos de las personas y las comu-

<sup>8</sup> Ver, GRI - Home (globalreporting.org).

## Marco ético de los datos del Sistema Estadístico Nacional

nidades. En este sentido, es fundamental contar con una gestión robusta del conflicto de intereses.

**Transparencia:** cada parte del sistema tiene la obligación de dar cuenta de sus acciones y resultados ante las entidades que hacen parte del SEN. Para ello, es fundamental, que cada instancia que haga parte del sistema tenga un plan de acción en relación con su sistema de gestión de los datos desde una perspectiva ética que se pueda discutir y hacer seguimiento.

**Interdisciplinariedad:** la complejidad en la gestión de los datos requiere que, en el proceso de seguimiento y evaluación, participen tanto profesionales de distintas disciplinas, expertos en el manejo de los datos,

como representantes de organizaciones de la sociedad civil que estén concernidas por el manejo de los datos de las entidades que hacen parte del SEN.

**Agilidad:** la oportunidad de la información es parte fundamental de su calidad. Por este motivo, el seguimiento que se haga debe ser ágil, de tal manera que se puedan hacer los correctivos en el manejo de los datos en el momento adecuado.

*El Sistema de evaluación de la gestión ética de los datos* estaría compuesto por tres grandes componentes: las entidades que hacen parte del SEN y que producen y consumen sus datos; el grupo de trabajo del SETE, y el CAD.

**Gráfica 3.** Estructura del sistema



Fuente: elaboración propia.

**a. Entidades del SEN:** ellas tendrían la obligación de tener un sistema de gestión de los datos que tenga una orientación ética. Para ello, las entidades deben contar con un marco ético que contenga principios similares o que sean homologables a los establecidos en el marco ético del SEN. Además de ello, deben contar con un sistema de evaluación y de seguimiento que garantice que en la gestión de los datos se están cumpliendo con los principios y los valores establecidos en su marco ético. Como se verá más adelante, las entidades adscritas al SEN deberán hacer un autodiagnóstico sobre la gestión ética de sus datos.

**b. El grupo de trabajo SETE:** esta unidad adscrita al DANE, tendrá las siguientes funciones:

- Evaluar a partir del autodiagnóstico en que grado las entidades cuentan con sistema de gestión de datos que

responde de manera efectiva con los principios establecidos en el marco ético del SEN. Para ello, verificarán la evidencia que las entidades entreguen como sustento de su autodiagnóstico.

- Acompañar a las entidades en el diseño y la implementación de sistemas de gestión de datos que cumplan con el marco ético.

Para que el SETE pueda cumplir de manera apropiada sus funciones, deben participar las cinco personas que actualmente lo conforman y al menos tres funcionarios de otras entidades que hacen parte del SEN para así garantizar una participación diversa y ampliar las capacidades de evaluación del grupo de trabajo.

**c. CAD:** la función central del CAD es dar el aval acorde con la evaluación que haga el SETE.

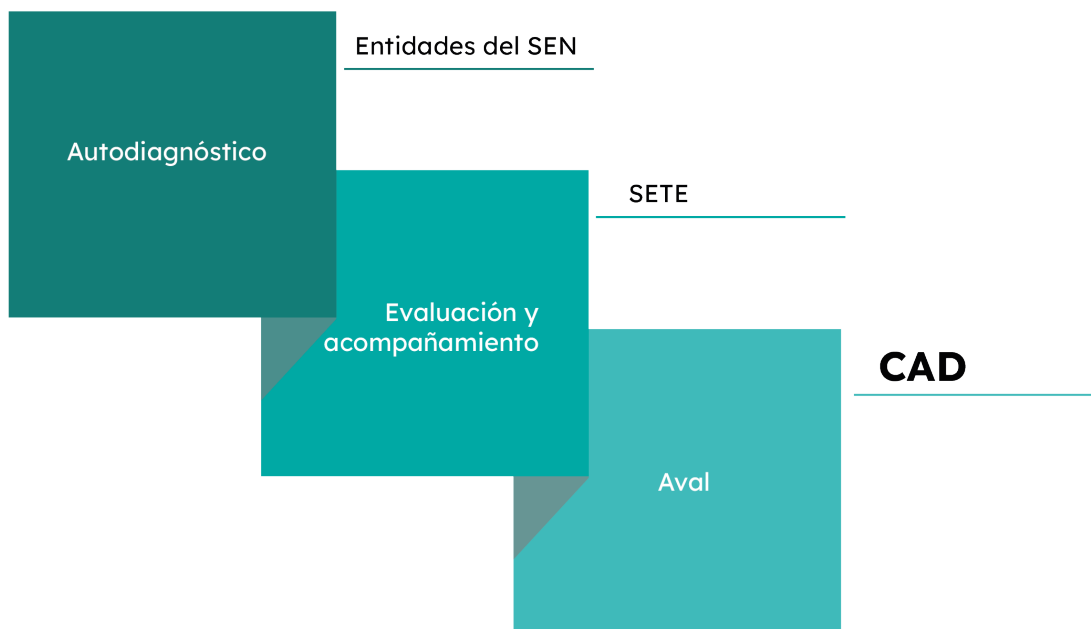
### 3.3. Proceso

Antes de presentar el proceso, y para que este sea más ágil, las entidades que sean evaluadas podrán presentar como insumos las certificaciones nacionales e internacionales que están hayan podido tener sobre sus procesos de gestión de los datos. Estas certificaciones se pueden llegar a homologar por el SETE y el CAD en la medida que en

ellas se tengan en cuenta principios similares a los expuestos en este marco y se lleven a cabo procedimientos de evaluación que sean independientes, transparentes y tengan en cuenta las normativas nacionales.

En líneas generales y teniendo en cuenta lo dicho anteriormente, el proceso tendría tres fases: autodiagnóstico; evaluación y acompañamiento, y aval.

**Gráfica 4.** Proceso del sistema



Fuente: elaboración propia.

**Autodiagnóstico:** en esta fase las entidades evalúan su proceso de gestión de datos teniendo en cuenta que este cumpla con los principios aquí señalados. Para este ejercicio las entidades tendrán a su disposición una guía elaborada por el grupo de trabajo SETE y podrán recibir una capacitación de este mismo grupo.

**Evaluación y acompañamiento:** en esta fase el grupo SETE hará una evaluación preliminar de la entidad, a partir del autodiagnóstico. Para este efecto, el grupo verificará

las evidencias que presente la institución y hará un análisis de fortalezas y debilidades.

**Aval:** el CAD, a partir de la evaluación preliminar, hará una evaluación definitiva de la institución y será la encargada oficial de establecer si la entidad tiene procesos de gestión de datos que cumplen con los principios establecidos en el marco ético.

### 3.4. Pasos a seguir

Para que esta propuesta se pueda poner en marcha se proponen las siguientes actividades:

- Discutir la propuesta con las principales entidades que hacen parte del SEN y hacer los ajustes respectivos. Para ello, se proponen hacer al menos dos grupos focales, cada uno con una duración de tres horas, en los que se discutan los valores y el diseño institucional.
- Elaborar una guía de autodiagnóstico de la gestión ética de los datos dirigida a las instituciones que hacen parte del SEN.
- Elaborar una guía de evaluación de la gestión ética.
- Elaborar un marco jurídico que recoja la propuesta.
- Establecer una nueva conformación de las instancias del SETE.
- Diseñar e implementar un programa de capacitación sobre gestión ética de los datos.
- Diseñar e implementar de una estrategia de comunicaciones en gestión ética para el SEN.

# Anexo A.

## Descripción de las fases del ciclo de vida de los datos

La siguiente tabla ha sido tomada directamente del Plan Nacional de Infraestructura de Datos (MinTIC et. Al, 2021) y en ella se describe cada fase del ciclo de vida de los datos:

Fase del ciclo de vida	Descripción
<b>Crear y obtener</b>	<p>Estos datos pueden tener múltiples formatos (por ejemplo: PDF, imagen, documento de Word), así como también de fuentes de datos estructurado y no estructurados o fuentes secundarias como las redes sociales.</p> <p>Los datos suelen ser creados por las organizaciones usualmente de las siguientes formas:</p> <p><b>Adquisición de datos:</b> adquisición de datos ya existentes que se han producido fuera de la organización.</p> <p><b>Entrada de datos:</b> entrada manual de nuevos datos por parte del personal dentro de la organización.</p> <p><b>Captura de datos:</b> captura de datos generados por dispositivos utilizados en varios procesos de la organización.</p> <p>El modelado y el diseño de datos es de suma importancia en esta fase, pues determina que información es la que se determina utilizar.</p>

<p><b>Procesar</b></p>	<p><b>Limpieza de datos:</b> en la que un conjunto de datos se limpia y se transforma de su forma sin procesar a algo más accesible y utilizable. Esto también se conoce corrección de datos.</p> <p><b>Compresión de datos:</b> en la que los datos se transforman en un formato que se puede almacenar de manera más eficiente.</p> <p><b>Cifrado de datos:</b> en el que los datos se traducen a otra forma de código para protegerlos de problemas de privacidad.</p> <p>Incluso el simple hecho de tomar un formulario impreso y digitalizarlo puede considerarse una forma de procesamiento de datos.</p> <p><b>Calidad de los datos:</b> determina que datos son utilizables aplicando las políticas y los controles establecidas en el diseño y aplicándolas en la obtención y el procesamiento de datos principalmente.</p>
<p><b>Almacenar</b></p>	<p>El almacenamiento y las operaciones incluyen el diseño, la implementación y el soporte de los datos almacenados para maximizar su valor, utilizando tanto formatos como repositorios que busquen el equilibrio entre disponibilidad y coste de almacenamiento, según los distintos escenarios de consulta, por ejemplo, diferenciando datos de alta demanda o disponibilidad de otros que no se consultan frecuentemente.</p> <p>Comúnmente el almacenamiento se realiza mediante la creación de bases de datos o conjuntos de datos. Estos conjuntos de datos pueden almacenarse en la nube, en servidores on premise o utilizando otras formas de almacenamiento como: discos duros magnéticos o de estado sólido, memorias o cintas magnéticas, entre otros.</p>
<p><b>Transferir y compartir</b></p>	<p>A medida que las organizaciones requieren de datos de fuentes secundarias, generados por otros actores del ecosistema de datos, útiles para la toma de decisiones, la planificación, la optimización de operaciones, entre otros, existe una mayor presión para compartir también datos generados. Los datos compartidos pueden ayudar a mejorar las decisiones, ya que las organizaciones pueden obtener una vista más completa de los impactos que sus decisiones han tenido en base a las contribuciones de nuevos conjuntos de datos de una variedad más amplia de fuentes, tanto internas como externas. Esta capacidad de compartición de datos debe estar soportada en una capa de interoperabilidad y haciendo uso de un estándar de lenguaje común de intercambio.</p>

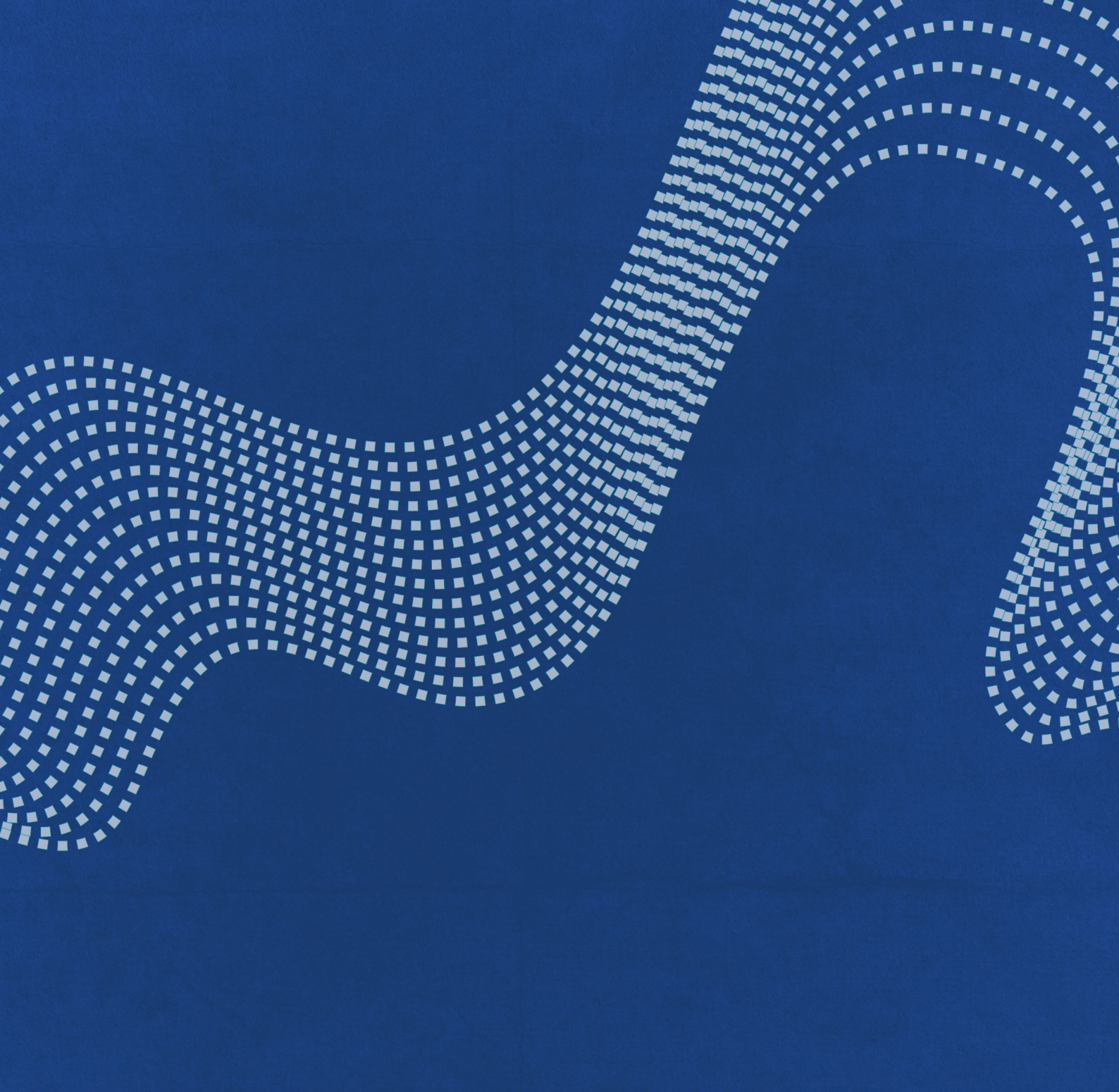
## Marco ético de los datos del Sistema Estadístico Nacional

<b>Analizar y usar</b>	<p>Durante la fase de uso del ciclo de vida de los datos, los datos se utilizan para respaldar las actividades de la organización. Los datos se pueden ver, procesar, modificar y guardar. Se debe mantener un registro de auditoría para todos los datos críticos, con el propósito de garantizar que todas las modificaciones que se realicen a los datos sean completamente rastreables y auditables. Los datos también pueden estar disponibles para exponer a otros actores del ecosistema de datos que se encuentren fuera de la organización.</p> <p>El análisis de datos se refiere a procesos que intentan obtener información significativa a partir de datos sin procesar. Los analistas y científicos de datos utilizan diferentes herramientas y estrategias para realizar estos análisis. Algunos de los métodos más utilizados incluyen modelado estadístico, algoritmos, inteligencia artificial, minería de datos y aprendizaje automático.</p>
<b>Archivar y preservar</b>	<p>El archivado de datos hace referencia a la copia de datos en un entorno donde se almacenan en caso de que se necesiten nuevamente en un ambiente de producción, asimismo también incluye la eliminación de estos datos de todos los entornos de producción activos.</p> <p>Un archivo de datos es simplemente un lugar donde se almacenan los datos, pero donde no se realiza mantenimiento o uso general. Si es necesario, los datos se pueden restaurar a un entorno en el que se puedan utilizar.</p> <p>Dado que el volumen de datos archivados crece inevitablemente, si bien es posible que desee guardar todos los datos de manera indefinida, los costos de almacenamiento pueden incentivar la destrucción de los datos que ya no se requieren. Por otro lado, es posible que se disponga de datos de uso limitado a una ventana de tiempo o a hasta /durante la ocurrencia de un suceso. La destrucción o la depuración de datos es la eliminación de cada copia de un elemento de datos de una organización.</p>
<b>Reutilizar</b>	<p>Reutilizar significa usar datos que originalmente fueron recopilados para otro fin o propósito. La reutilización de datos también se puede llamar análisis secundario.</p>



# Bibliografía

- Beauchamp, Tom, Walters, Leroy. (1999). Contemporary Issues in Bioethics. Londres, Wadsworth Publishing Company.
- Buenadicha, César, Galdon, Gemma, Hermosilla, María Paz, Loewe, Daniel y Pombo Cristina. (2019). La gestión ética de los datos. Banco Interamericano de Desarrollo (BID).
- Comisión Económica de las Naciones Unidas para Europa [UNECE]. (2022). Core Values of Official Statistics. United Nations. <https://unece.org/statistics/documents/2022/06/working-documents/core-values-official-statistics>
- Congreso de la República de Colombia (2023). Ley 2335 de 2023 “Por la cual se expiden disposiciones sobre las estadísticas oficiales en el país”. <https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=221910>
- Davis, Kord con Paterson Doug. (2012). Ethics of big data. Cambridge: O’Reilly.
- Departamento Administrativo Nacional de Estadística [DANE]. (2023). Resolución 2259 del 2023. Por la cual se regula la estructura, el funcionamiento y la secretaría técnica del Comité de Administración de Datos (CAD).
- García-Marzá, Domingo (2004). Ética empresarial: del diálogo a la confianza. Trotta, 2004, pp. 261-284.
- Gobierno de Colombia (2021). Marco Ético para la Inteligencia Artificial en Colombia. <https://inteligenciaartificial.gov.co/marco-eticos/>
- Hasselbalch, Gry, Tranberg Pernille. (2016). Data Ethics: The New Competitive Advantage. Aka Print.
- Hansson, Sven Oven, (Editor) (2017). The ethics of technology: methods and approaches. New York: Rowman & Littlefield International, pp. 115-129.
- MinTIC. (2008). Guía de Estándares de Interoperabilidad Gobierno en Línea. (S.F).
- ISO 25912, Software y datos.
- MinTIC, DNP y Presidencia de la República de Colombia (2021). Plan Nacional de Infraestructura de Datos.
- Naciones Unidas. (2022). Enfoque para la programación basado en los derechos humanos. <https://unsdg.un.org/es/2030-agenda/universal-values/human-rights-based-approach>
- Unión Europea. (2016). Reglamento General de Protección de Datos para la Unión Europea, (RGPD).
- Unión Europea (2022) Regulation of the european parliament and of the council on harmonised rules on fair access to and use of data (Data Act). <https://digital-strategy.ec.europa.eu/en/news/data-act-commission-proposes-measures-fair-and-innovative-data-economy>



@DANE\_Colombia



/DANEColombia



/DANEColombia



@DANEColombia

[www.sen.gov.co](http://www.sen.gov.co)