

Lineamientos para la anonimización de información no estructurada en documentos de archivo en entidades públicas



Archivo General
de la Nación



ARCHIVO GENERAL DE LA NACIÓN JORGE PALACIOS PRECIADO - COLOMBIA

Establecimiento público adscrito al Ministerio de las Culturas, las Artes y los Saberes

Director General

Francisco Javier Flórez Bolívar

Secretaria General

Yanina Rosa Arrieta Leottau

Subdirector de Política y Normativa Archivística

William Escobar Quintero

Subdirector de Inspección, Vigilancia y Control

Ignacio Manuel Epinayu Pushaina

Subdirectora de Gestión del Patrimonio Documental

Rosario Arias Callejas

Subdirector del Sistema Nacional de Archivos

Jorge Luis Turizo Correa

Subdirectora de Transformación Digital e Innovación Archivística

Yira Marcela Montilla

Subdirectora de Archivos de Entidades Liquidadas

Gloribel Rodríguez Carrasco

Subdirectora de Mercadeo y Operación Servicios Archivísticos

Derly Isaac Cañaveral

Jefe Oficina Asesora Jurídica

Margarita Cecilia Arrieta Ramírez

Jefe Oficina Asesora de Planeación

Raul Salas Cassiani

Jefe Oficina de Control Interno

Gerardo Ariza Ariza

Asesores de Dirección

Karent Rocío Cabezas Rosero

María Camila Díaz Casas

Arturo Daniel López Coba

Mariela Lugo Maldonado

Investigación

Heidy Elieth Balanta

Fernando Augusto Ibarra Martínez

John Alexander González Flores

Diego Armando Ortiz Saza

Jessica Mayerly Molina Buitrago

German Ricardo Cubides Pinto

Edición

Maria Paula Díaz Castro

Diseño y diagramación

Diego Alejandro Rodríguez Gracia

Archivo General de la Nación de Colombia

Carrera 6 No 6 – 91

Teléfono: 601 328 2888

E-mail: contacto@archivogeneral.gov.co

Página web: www.archivogeneral.gov.co

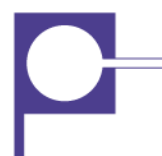
Bogotá D.C., Colombia - 2026

Contenido

Objetivo.....	4
Alcance.....	4
Destinatarios	4
Presentación	6
Interpretación de los presentes lineamientos.....	7
¿Por qué anonimizar?	8
PRIMERA PARTE.....	10
1. Conceptos fundamentales para la anonimización.....	11
2. Protección de datos personales y la función archivista	15
2.1 Principios del régimen de protección de datos personales y su relación con la función archivística.....	17
3. Algunos supuestos en los que las entidades públicas deben anonimizar	19
4. Principios aplicables al proceso de anonimización	23
5. Marco legal de la anonimización en Colombia.....	25
5.1 Ley 1581 de 2012. Régimen general de la protección de datos personales en Colombia.....	25
5.2 Ley 1712 de 2014. Régimen general de transparencia y acceso a la información pública.....	26
5.3 Legislación complementaria.....	27
6. La información no estructurada en los documentos de archivo.....	30
6.1 Formatos no estructurados que pueden contener datos personales o información sensible.....	31
6.2 Datos específicos dentro de formatos no estructurados.....	32
6.3 Información no estructurada que se utiliza en las entidades públicas.....	33
7. Tipos de datos personales sujetos a anonimización, identificadores directos e indirectos	35
7.1 Categorías de datos personales según su nivel de protección	39
8. Sujetos, espacios y contextos objeto de protección mediante procesos de anonimización.....	40
8.1 Categoría de sujetos protegidos.....	40
8.2 Espacios, lugares y hechos objeto de protección	44
9. Técnicas de anonimización para información no estructurada	47
9.1 Supresión o eliminación directa.....	48
9.2 Sustitución por etiquetas o codificación.....	49
9.3 Generalización.....	49
9.4 Disociación irreversible	51



9.5	Distorsión técnica en formatos audiovisuales	52
9.6	Seudonimización controlada.....	53
9.7	Acceso parcial o mediado.....	54
9.8	Combinación de técnicas	55
10.	Riesgos de reidentificación	55
10.1	Modalidades técnicas de riesgo de reidentificación	56
10.2	Inteligencia artificial y los retos de reidentificación.....	58
10.3	Consecuencias de una reidentificación.....	59
11.	Enfoque de anonimización según el ciclo de vida documental.....	61
11.1	Lineamiento de anonimización en archivos de gestión.....	63
11.2	Lineamiento de anonimización en archivo central.....	64
12.	Lineamiento de anonimización en archivo histórico	65
13.	Lineamientos metodológicos en el proceso de anonimización	67
13.1	Identificación de información susceptible de anonimización.....	68
13.2	Determinación del régimen jurídico aplicable	69
13.3	Intervención técnica sobre el documento de archivo	71
13.4	Intervención en formato físico.....	71
13.5	Intervención en formato digital textual.....	72
13.6	Intervención en imágenes y fotografías	74
13.7	Intervención en formato audio	75
13.8	Intervención en formato video	76
13.9	Intervención en expedientes híbridos	77
13.10	Intervención en elementos accesorios y capas técnicas no visibles.....	78
13.11	Verificación técnica final y validación previa a la divulgación	79
13.12	Medidas compensatorias cuando la anonimización no sea viable.....	80
13.13	Mecanismos institucionales de acceso controlado cuando la anonimización no sea viable.....	82
14.	Lineamientos para la selección de técnicas de anonimización en información no estructurada	85
15.	Análisis de riesgo contextual y probabilidad de reidentificación	87
16.	Estructura institucional y responsabilidades	88
17.	Bibliografía.....	94



Objetivo

Establecer lineamientos generales, jurídicos, técnicos conceptuales y metodológicos para la anonimización de información contenida en documentos de archivo que contienen datos personales, con el fin de proteger a las personas, mitigar riesgos de reidentificación, revictimización y armonizar el régimen de protección de datos personales (autodeterminación informativa) con el derecho de acceso a la información pública, garantizando así la transparencia, la seguridad jurídica y la calidad de la información publicada.

Alcance

El presente documento aplica a los procesos de identificación, evaluación y aplicación de técnicas de anonimización y seudonimización sobre documentos de archivo que contengan datos personales, información reservada o clasificada, o cualquier otro tipo de información cuya divulgación pueda afectar derechos fundamentales, la seguridad nacional o los intereses legítimos de terceros, y que sea objeto de descripción, acceso, publicación, divulgación o entrega a terceros, ya sea de manera proactiva, en virtud de solicitudes de acceso a la información pública.

Destinatarios

El documento está dirigido a:

- **Las entidades públicas** de todos los niveles y ramas del poder público¹
- Las entidades privadas que cumplen funciones públicas y los demás organismos regulados por la Ley 594 de 2000
- Los servidores públicos y contratistas que intervienen en:
 - Gestión documental y archivística
 - Transparencia y acceso a la información pública
 - Protección de datos personales (autodeterminación informativa)
 - Seguridad de la información
 - Tecnologías de la información
 - Atención al ciudadano y relatoría, cuando aplique

¹ Las entidades públicas implementarán los lineamientos establecidos en el presente documento en aquella información que sea de su propiedad, de la cual adquiera la calidad de tal o de que, sin ser de su propiedad, el titular lo autorice para tal fin.

- Oficiales de Protección de Datos, responsables y encargados del tratamiento de datos
- Áreas jurídicas y de control interno que apoyan la toma de decisiones sobre publicación y divulgación de información

De manera opcional, los presentes lineamientos podrán ser considerados por organizaciones sociales que administren archivos o gestionen información documental, en la medida en que su contenido pueda resultar útil para fortalecer sus prácticas de anonimización y protección de datos.

Presentación

Los presentes lineamientos constituyen un instrumento técnico elaborado por el Archivo General de la Nación Jorge Palacios Preciado con el propósito de orientar a las entidades públicas en la aplicación de procesos de anonimización sobre información no estructurada contenida en documentos de archivo. Su elaboración responde a la necesidad de armonizar derechos fundamentales que con frecuencia se tensionan en la práctica institucional. El derecho de acceso a la información pública y el derecho a la protección de datos personales y a la intimidad constituyen la tensión más frecuente y visible. Sin embargo, la práctica archivística revela que esta tensión raramente es binaria: en un mismo documento pueden coexistir el derecho a la verdad y a la memoria histórica colectiva, que exige que ciertos responsables sean identificables; el derecho a la seguridad personal y a la vida de quienes aparecen en los documentos y cuya identificación puede representar un riesgo real incluso décadas después de los hechos; el derecho a la honra y al buen nombre de personas vinculadas a información que no han podido controvertir; el derecho a la reparación, que puede requerir que determinada información sea pública y atribuible; y el derecho al olvido, reconocido por la Corte Constitucional como una dimensión del habeas data, que puede tensionarse con el deber de memoria histórica.

La anonimización es el mecanismo técnico y jurídico que hace posible navegar entre esas tensiones. Permite abrir el archivo sin desproteger a quienes aparecen en él, y habilita el cumplimiento simultáneo de obligaciones que de otro modo resultarían contradictorias. No obstante, su aplicación no es automática ni neutral: exige siempre un ejercicio de ponderación cuidadoso, contextualizado y jurídicamente fundamentado, que considere el tipo de documento, el soporte, el contenido, los sujetos involucrados y el uso previsto de la información.

El documento se organiza en dos partes claramente diferenciadas. La primera desarrolla los fundamentos conceptuales, jurídicos y técnicos de la anonimización. Parte de las definiciones esenciales y de la relación entre la función archivística y el régimen de protección de datos personales, para luego recorrer el marco normativo colombiano aplicable y los principios que deben orientar todo proceso de anonimización. A partir de allí, el documento caracteriza los formatos no estructurados presentes en los acervos documentales de las entidades públicas, identifica las categorías de datos personales sujetos a protección, describe los sujetos y contextos que demandan intervención prioritaria, y presenta el catálogo de técnicas disponibles, desde la supresión directa hasta la distorsión en formatos audiovisuales, junto con los riesgos de reidentificación que deben gestionarse, incluidos los que plantea el uso de inteligencia artificial.

La segunda parte traduce esos fundamentos en lineamientos operativos concretos. Aborda la anonimización desde el ciclo de vida documental, estableciendo criterios diferenciados para el archivo de gestión, el archivo central y el archivo histórico. Desarrolla luego una metodología paso a paso para la intervención técnica sobre los distintos soportes: físico, digital, sonoro, audiovisual e híbrido, e incluye orientaciones para la selección de técnicas, el análisis de riesgo contextual y las medidas aplicables cuando la anonimización no resulte viable. La segunda parte cierra con los lineamientos sobre estructura institucional y responsabilidades, definiendo los roles que deben asumir las entidades en la gobernanza de sus procesos de anonimización.

En conjunto el documento ofrece a las entidades públicas una hoja de ruta integral que articula el derecho, la archivística y la técnica, con el propósito de garantizar que la apertura de la información pública se realice siempre de manera responsable, proporcional y respetuosa de los derechos fundamentales.

Interpretación de los presentes lineamientos

Los presentes lineamientos constituyen una herramienta técnica y orientadora elaborada por el Archivo General de la Nación Jorge Palacios Preciado, en ejercicio de sus funciones como ente rector del Sistema Nacional de Archivos, con el propósito de apoyar a las entidades públicas en la estructuración y fortalecimiento de sus procesos de anonimización de documentos de archivo. En virtud del principio de unidad normativa y descentralización operativa que rige el Sistema Nacional de Archivos, estos lineamientos están dirigidos a todas las entidades que lo integran, sin perjuicio de las competencias legales propias de cada una y de las particularidades normativas que aplican según la rama del poder público o el nivel territorial al que pertenezcan.

Su aplicación deberá entenderse como un punto de partida para el análisis institucional, que cada entidad deberá adaptar a su naturaleza jurídica, funciones misionales, nivel de riesgo, capacidades técnicas y el contexto documental específico. En consecuencia, corresponde a cada sujeto obligado evaluar sus necesidades particulares, definir sus propios procedimientos internos y adoptar las medidas que resulten proporcionales y adecuadas conforme al marco constitucional y legal vigente.

Estos lineamientos no crean obligaciones nuevas ni modifican las existentes; su finalidad es ofrecer criterios técnicos, metodológicos y archivísticos que sirvan de referencia para la toma de decisiones responsables en materia de anonimización.

¿Por qué anonimizar?

La anonimización es el resultado de un proceso aplicado a datos personales con el propósito de impedir de manera irreversible la identificación del titular. Para que este estándar se cumpla, la información debe ser tratada de modo que no sea posible identificar a una persona mediante el uso de todos los medios razonablemente disponibles, evaluando los riesgos de singularización, vinculabilidad e inferencia (Article 29 Working Party, 2014)².

La anonimización garantiza el derecho de acceso a la información pública y el derecho fundamental a la protección de datos personales. Es una herramienta que permite ejercer ambos derechos de forma simultánea y responsable. Un ciudadano que confía en que sus datos serán tratados con responsabilidad participa más activamente, denuncia con mayor libertad y utiliza los canales institucionales sin temor. Esa confianza es, en sí misma, un activo democrático que fortalece la relación entre el Estado y la sociedad.

En el contexto colombiano, la anonimización abre posibilidades concretas y valiosas. Permite avanzar en la reconstrucción de la memoria histórica y dar respuesta a sentencias en firme, contribuye a la seguridad del Estado y fortalece el régimen democrático. Hace posible que investigadores, historiadores y ciudadanos accedan a hechos, contextos y patrones estructurales relevantes para los procesos de justicia transicional, sin exponer innecesariamente la intimidad de las personas involucradas. En pocas palabras, abre el archivo sin desproteger a quienes aparecen en él.

El acceso a la información y su divulgación responsable no son opciones institucionales, sino principios fundamentales del Estado Social de Derecho, reconocidos tanto en el ordenamiento jurídico colombiano, como en el derecho internacional de los derechos humanos. El *Conjunto de Principios para la Protección y la Promoción de los Derechos Humanos mediante la Lucha contra la Impunidad* (2005), conocido como “Principios Orentlicher”³, establece que toda persona tiene el derecho inalienable a conocer la verdad acerca de los acontecimientos sucedidos y las circunstancias que llevaron a las graves violaciones de los derechos humanos. Este derecho a la verdad no se agota en el ámbito individual: es también un derecho colectivo que pertenece a la sociedad en su conjunto, y que se materializa, entre otros mecanismos, a través del acceso a

2 Article 29 Working Party. (2014). Opinion 05/2014 on Anonymisation Techniques (WP216). Bruselas: Comisión Europea.

3 Naciones Unidas, Comisión de Derechos Humanos. (2005). Conjunto de Principios actualizado para la Protección y la Promoción de los Derechos Humanos mediante la Lucha contra la Impunidad (Principios Orentlicher). Documento E/CN.4/2005/102/Add.1, 8 de febrero de 2005. Disponible en: <https://undocs.org/E/CN.4/2005/102/Add.1>

los archivos del Estado. En consecuencia, los lineamientos de anonimización consideran, además del marco normativo nacional de protección de datos y de archivos, también el conjunto de principios internacionales que imponen al Estado la obligación simultánea de abrir sus archivos y proteger a quienes en ellos aparecen.

En el marco de la desclasificación, la anonimización es una práctica que articula, de forma, virtuosa consideraciones jurídicas, éticas, técnicas, archivísticas e históricas. Sus decisiones inciden directamente en el valor probatorio, interpretativo y testimonial de los documentos, y por eso constituye un ejercicio de rigor institucional al servicio de la verdad y la justicia.

Por tanto, el Estado colombiano avanza hacia esquemas de datos abiertos, interoperabilidad, analítica avanzada e inteligencia artificial. La anonimización es la base que hace ese futuro posible: sin ella la apertura de datos podría generar riesgos; con ella, se convierte en una oportunidad real para una gestión pública más inteligente, más justa y confiable.

Anonimizar es, en definitiva, una garantía democrática, una expresión de respeto hacia los ciudadanos y una obligación ética del Estado. Es la forma en que Colombia puede construir un Estado verdaderamente abierto, que protege a las personas mientras les sirve de la mejor manera.

PRIMERA PARTE

Marco conceptual y fundamentos de la anonimización

- 1 Conceptos fundamentales para la anonimización
- 2 Protección de datos personales y la función archivística
- 3 Algunos supuestos en los que las entidades públicas deben anonimizar
- 4 Marco legal de la anonimización en Colombia
- 5 Principios aplicables al proceso de anonimización
- 6 Documentos de archivo como información no estructurada
- 7 Sujetos, espacios y contextos, objeto de protección mediante procesos de anonimización
- 8 Tipos de datos personales sujetos a anonimización, identificadores directos e indirectos
- 9 Técnicas de anonimización para información no estructurada
- 10 Riesgos de reidentificación

1. Conceptos fundamentales para la anonimización

Las siguientes definiciones constituyen el marco conceptual requerido para la aplicación de los lineamientos de anonimización de datos e información contenida en documentos de archivo, independientemente del soporte en el cual se encuentren.

- **Administración de archivos:** Conjunto de estrategias organizacionales dirigidas a la planeación, dirección y control de los recursos físicos, técnicos, tecnológicos, financieros y del talento humano, para el eficiente funcionamiento de los archivos (Acuerdo 001 de 2024, AGN)
- **Acceso a documentos de archivo:** Derecho de los ciudadanos y atribución de las autoridades a consultar la información que conservan los archivos de las instituciones u organizaciones que desarrollan funciones públicas, en los términos consagrados por la ley; dicho ejercicio estará limitado cuando se trate de información sujeta a reserva, previa autorización por la Constitución y la ley (*ibid.*)
- **Acervo documental:** Conjunto de documentos de un archivo, conservados por su valor sustantivo, histórico o cultural (*ibid.*)
- **Anonimización:** "Proceso mediante el cual se modifica la información de identificación personal de forma irreversible de tal manera que no se pueda identificar, directa o indirectamente, al titular, ya sea por sus propios medios o en colaboración con un tercero" (Guía de Anonimización de Datos Estructurados-AGN)
- **Archivo⁴:** Conjunto de documentos, sea cual fuere su fecha, forma y soporte material, acumulados en un proceso natural por una persona o entidad pública o privada, en el transcurso de su gestión, conservados respetando aquel orden para servir como testimonio e información a la persona o institución que los produce y a los ciudadanos, o como fuentes de la historia. También se puede entender como la institución que está al servicio de la

4 De conformidad con el Principio 14 del Conjunto de Principios para la Protección y la Promoción de los Derechos Humanos mediante la Lucha contra la Impunidad (Principios Orentlicher, ONU 2005), la palabra "archivos" se refiere a colecciones de documentos relativos a violaciones de los derechos humanos y el derecho humanitario de fuentes que incluyen: a) organismos gubernamentales nacionales, en particular los que hayan desempeñado una función importante en relación con las violaciones de los derechos humanos; b) organismos locales, tales como comisarías de policía, que hayan participado en violaciones de los derechos humanos; c) organismos estatales, incluida la oficina del fiscal y el poder judicial, que participan en la protección de los derechos humanos; y d) materiales reunidos por las comisiones de la verdad y otros órganos de investigación.

gestión administrativa, la información, la investigación y la cultura (Acuerdo 001 de 2024, AGN)

- **Archivo central:** Unidad administrativa que coordina y controla el funcionamiento de los archivos de gestión y reúne los documentos transferidos por los mismos una vez finalizado su trámite y cuando su consulta es constante (*ibid.*)
- **Archivo de gestión:** Archivo de la oficina productora que reúne su documentación en trámite, sometida a continua utilización y consulta administrativa (*ibid.*)
- **Archivo electrónico:** Conjunto de documentos electrónicos producidos y tratados archivísticamente, siguiendo la estructura-orgánico funcional del productor, acumulados en un proceso natural por una persona o institución pública o privada, en el transcurso de su gestión (*ibid.*)
- **Autenticidad:** Característica técnica que permite identificar al autor de un mensaje de datos, el cual es conservado en condiciones que permitan garantizar su integridad, para preservar la seguridad de la información que busca asegurar su validez en tiempo, forma y distribución. Así mismo, garantiza el origen de la información, validando el emisor para evitar suplantación de identidades. Que pueda demostrarse que el documento es lo que afirma ser, que ha sido creado o enviado por la persona que afirma haberlo creado o enviado, y que ha sido creado o enviado en el momento que se afirma (*ibid.*)
- **Datos estructurados:** Son datos que están organizados en forma de tablas o matrices, como archivos de Excel o bases de datos relacionales, donde se definen las relaciones entre ellos (Guía de Anonimización de Datos Estructurados-AGN)
- **Datos no estructurados:** Datos cuya organización y presentación no está guiada por ningún modelo o esquema predefinido. En esta categoría se incluyen "imágenes, textos, audios, contenidos de redes sociales, videos" o documentos en formato PDF y Word (*ibid.*)
- **Dato personal:** Cualquier información vinculada o que pueda asociarse a una o varias personas naturales determinadas o determinables (literal c) artículo 3, Ley 1581 de 2012)
- **Dato privado:** Es el dato que por su naturaleza íntima o reservada solo es relevante para el titular (literal h), artículo 3, Ley 1266 de 2008)
- **Dato público:** Es el dato que no sea semiprivado, privado o sensible. Son considerados datos públicos, entre otros, los datos relativos al estado civil de las personas, a su profesión u oficio y a su calidad de comerciante o de servidor público. Por su naturaleza, los datos públicos pueden estar contenidos, entre otros, en registros públicos, documentos públicos, gacetas y boletines oficiales y, sentencias judiciales, debidamente ejecutoriadas, que no estén sometidas a reserva (artículo 2.2.2.25.1.3, del Decreto 1074 de 2015)
- **Dato semiprivado:** Es semiprivado el dato que no tiene naturaleza íntima, reservada, ni pública y cuyo conocimiento o divulgación puede interesar no solo a su titular, sino a cierto

sector o grupo de personas o a la sociedad en general, como el dato financiero y crediticio de actividad comercial o de servicios (literal g), artículo 3, Ley 1266 de 2008)

- **Datos sensibles:** Se entiende por datos sensibles aquellos que afectan la intimidad del titular o cuyo uso indebido puede generar su discriminación, tales como aquellos que revelen el origen racial o étnico, la orientación política, las convicciones religiosas o filosóficas, la pertenencia a sindicatos, organizaciones sociales, de derechos humanos o que promueva intereses de cualquier partido político o que garanticen los derechos y garantías de partidos políticos de oposición, así como los datos relativos a la salud, a la vida sexual, y los datos biométricos (Artículo 5, de la Ley 1581 de 2012)
- **Desidentificación:** Consiste en la eliminación de identificadores directos (por ejemplo, nombre, dirección, número de documento nacional de identidad) que identifican directamente a un individuo. Es solo el primer paso de la anonimización; un conjunto de datos desidentificado puede volver a identificarse fácilmente si se combina con otros datos (PDPC & AEPD, 2022)
- **Digitalización:** Técnica que permite la reproducción de información que se encuentra guardada de manera analógica (papel, video, cassettes, cinta, película, microfilm y otros) en una que sólo puede leerse o interpretarse por sistemas de información (Acuerdo 001 de 2024, AGN)
- **Documento electrónico de archivo:** Registro de información generada, producida o recibida o comunicada por medios electrónicos, que permanece almacenada electrónicamente durante todo su ciclo de vida, producida, por una persona o entidad en razón a sus actividades o funciones, que tiene valor administrativo, fiscal, legal, o valor científico, histórico, técnico o cultural y que debe ser tratada conforme a los principios y procesos archivísticos (*ibid.*)
- **Expediente digital o digitalizado:** Copia exacta de un expediente físico cuyos documentos originales, tradicionalmente impresos, son convertidos a formato electrónico mediante procesos de digitalización (*ibid.*)
- **Expediente electrónico:** Conjunto de documentos electrónicos correspondientes a un procedimiento administrativo, cualquiera que sea el tipo de información que contengan. El expediente electrónico deberá garantizar condiciones de autenticidad, integridad y disponibilidad (*ibid.*)
- **EXIF:** (*Exchangeable Image File Format*, que traduce Formato de Archivo de Imagen Intercambiable). Los archivos Exif son metadatos producidos por algunos formatos de imagen, que funcionan como indicadores de Información de Identificación Personal (PII, por sus siglas en inglés), ya que revelan datos contextuales ocultos como la geolocalización. La información contenida en estos formatos se produce mediante datos binarios, lo que hace que su identificación y eliminación sistemática (limpieza) sea un proceso complejo, que

requiere conocimientos amplios sobre dichos formatos para evitar vulnerar la privacidad de los individuos (Guía para la Anonimización de Datos e Información No Estructurada: Estándares y Lineamientos Técnicos- CNMH)

- **Formato electrónico:** Estructura técnica y lógica en la que se codifica y organiza la información de un documento, que determina la forma en que puede ser interpretada y procesada por un sistema informático o visualizada por el usuario (por ejemplo, PDF, DOCX, XLSX, TIFF, MP4) (Definición propia, AGN)
- **Índice de información clasificada y reservada:** Es el inventario de la información pública generada, obtenida, adquirida o controlada por el sujeto obligado, en calidad de tal, que ha sido calificada como clasificada o reservada (Artículo 2.8.5.2.1., Decreto 1080 de 2015)
- **Material gráfico:** Todos aquellos dibujos, croquis, mapas, planos, fotografías, ilustraciones, pictografías, códigos, prensa, entre otros (Acuerdo 001 de 2024, AGN)
- **Medio electrónico:** Mecanismo tecnológico, óptico, telemático, informático o similar, conocido o por conocerse, que permite producir, almacenar o transmitir documentos, datos o información (*ibid.*)
- **OCR (Optical Character Recognition: Reconocimiento Óptico de Caracteres):** tecnología que permite convertir imágenes de texto (como documentos escaneados o fotografías) en texto digital editable y procesable por sistemas informáticos (Adobe)
- **Programa de gestión documental (PGD):** Es el instrumento archivístico que permite a los sujetos obligados, formular y documentar a corto, mediano y largo plazo, el desarrollo sistemático de los procesos de gestión documental, encaminados a la planificación, procesamiento, manejo y organización de documentación producida y recibida, desde su origen hasta su destino final, para facilitar su uso, conservación y preservación (Acuerdo 001 de 2024, AGN)
- **Recepción de documentos:** Conjunto de operaciones de verificación y control que una institución debe realizar para la admisión de los documentos que le son remitidos por una persona natural o jurídica (*ibid.*)
- **Recuperación de documentos:** Acción y efecto de obtener, por medio de los instrumentos de consulta, los documentos requeridos (*ibid.*)
- **Reidentificación:** Se refiere a la identificación de individuos a partir de un conjunto de datos que previamente fue desidentificado o anonimizado (PDPC & AEPD, 2022)
- **Repositorios digitales:** Sistema de almacenamiento que tiene como objetivo conservar, preservar y almacenar los documentos electrónicos de archivo de acuerdo con los tiempos de retención establecidos en la Tabla de Retención Documental, mientras se encuentren bajo su custodia. Un repositorio digital de confianza es un sistema que almacena datos digitales de manera segura. Estos datos pueden incluir documentos, imágenes, videos, audio y otros tipos de archivos. Los repositorios digitales de confianza utilizan una variedad de

técnicas para proteger los datos de acceso no autorizado, pérdida o destrucción. Estas técnicas pueden incluir encriptación, controles de acceso y software de detección y prevención de hackeos (Acuerdo 001 de 2024, AGN)

- **Serie documental:** Documentos organizados de acuerdo con un sistema de archivo o conservados formando una unidad como resultado de la acumulación del mismo proceso archivístico, o de la misma actividad; que tienen una forma particular o como consecuencia de cualquier actividad derivada de su producción, recepción o utilización (*ibid.*)
- **Seudonimización:** Es la sustitución de un atributo (normalmente un atributo único) por otro valor en un registro (seudónimo). A diferencia de la anonimización, "el uso exclusivo de la seudonimización no garantiza un conjunto de datos anónimo", ya que permite revertir el proceso si se tiene la clave (Guía de Anonimización de Datos Estructurados-AGN)
- **Soporte documental:** Medios en los cuales se contiene la información, según los materiales empleados. Además de los archivos en papel existente los archivos audiovisuales, fotográficos, fílmicos, informáticos, orales y sonoros (Artículo 3, Ley 594 de 2000)
- **Soporte físico o analógico:** Medio material tangible en el cual se fija la información, tales como papel, microformas, cintas magnéticas u otros materiales físicos (Acuerdo 001 de 2024, AGN)
- **Subserie documental:** Conjunto de unidades documentales que forman parte de una serie, identificadas de forma separada de esta por su contenido y sus características específicas (*ibid.*)
- **Versión pública:** Copia de un documento de archivo en la que se suprime o protege la información clasificada y/o reservada, manteniendo el resto del contenido accesible conforme a la normativa vigente, sin alterar el documento original (Artículo 21, Ley 1712 de 2014)

2. Protección de datos personales y la función archivista⁵

Cuando se hace referencia a documentos, específicamente a información, ha de considerarse que la información contiene datos, personales, que son objeto de tratamiento, estos se encuentran sometidos al cumplimiento de la Ley 1581 de 2012. El fundamento normativo de los archivos y de los datos personales (*Habeas Data*) se encuentra en la Constitución Política de Colombia:

⁵ Extracto del capítulo "Habeas Data y Función Archivística", de libro *Documentos, mensajes de datos*. Arturo Daniel López Coba. Sin publicar.

Todas las personas tienen derecho a su intimidad personal y familiar y a su buen nombre, y el Estado debe respetarlos y hacerlos respetar. De igual modo, tienen derecho a conocer, actualizar y rectificar las informaciones que se hayan recogido sobre ellas en bancos de datos y en archivos de entidades públicas y privadas (artículo 15)

La jurisprudencia de la Corte Constitucional⁶, relacionada con el *habeas data* y archivos, expresa lo siguiente:

Para la Sala, la preocupación del ciudadano es muy importante, pues ciertamente los archivos contienen datos personales susceptibles de ser tratados y, por tanto, que requieren medidas de protección; sin embargo, la Sala observa que los archivos sí hacen parte del ámbito de aplicación de la ley, por las siguientes razones:

El artículo 3 del proyecto define las bases de datos de una manera muy amplia como el “[c]onjunto organizado de datos personales que sea objeto de Tratamiento”. El tratamiento, por su parte, es definido como “[c]ualquier operación o conjunto de operaciones sobre datos personales, tales como recolección, almacenamiento, uso, circulación o supresión”.

El proyecto no contiene una definición de archivo; sin embargo, éste es definido por la Real Academia de la Lengua como el “[c]onjunto ordenado de documentos que una persona, una sociedad, una institución, etc., producen en el ejercicio de sus funciones o actividades” o como el “[l]ugar donde se custodian uno o varios archivos.” Los archivos también son definidos por la Ley 594 de 2000 “por medio de la cual se dicta la Ley General de Archivos y se dictan otras disposiciones”, como el “[c]onjunto de documentos, sea cual fuere su fecha, forma y soporte material, acumulados en un proceso natural por una persona o entidad pública o privada, en el transcurso de su gestión, conservados respetando aquel orden para servir como testimonio e información a la persona o institución que los produce y a los ciudadanos, o como fuentes de la historia (artículo 3)

Y concluye la citada Corporación Judicial⁷:

De acuerdo con estas definiciones, los archivos -para efectos exclusivamente del proyecto-, en tanto son (i) depósitos ordenados de datos, incluidos datos personales, y (ii) suponen, como mínimo, que los datos han sido recolectados, almacenados y, eventualmente, usados –

6 Sentencia C-748 de octubre 06 de 2011. M.P. Jorge Ignacio Pretelt Chaljub. Fundamento jurídico 2.4.3.2.

7 Corte Constitucional. Sentencia C-748 de octubre 06 de 2011. M.P. Jorge Ignacio Pretelt Chaljub. Fundamento jurídico 2.4.3.2.

modalidades de tratamiento, son una especie de base de datos que contiene datos personales susceptibles de ser tratados y, en consecuencia, serán cobijados por la ley una vez entre en vigencia.

2.1 Principios del régimen de protección de datos personales y su relación con la función archivística

Descrito lo anterior, se puede afirmar que, se integran y complementan los principios del *Habeas Data* y la función archivística, para la protección de derechos y garantías fundamentales, como pasa a describirse:

(PRINCIPIOS) HABEAS DATA⁸ Ley 1581 de 2012. Artículo 4°. Principios para el tratamiento de datos personales. Corte Constitucional. Sentencia C-748 de 2011.	FUNCIÓN ARCHIVÍSTICA. Ley 594 de 2000. Artículo 4. Principios generales. Corte Constitucional. Sentencia T-656 de 2010.
(i) Principio de libertad	Ley 594 de 2000. Artículo 4. Principios generales [...] Responsabilidad
(ii) Principio de necesidad	Principio de orden original
(iii) Principio de veracidad o calidad	Ley 594 de 2000. Artículo 11
(iv) Principio de integridad en el manejo de datos	Principio de orden original. Ley 594 de 2000. Artículo 11
(v) Principio de finalidad	Ley 594 de 2000. Artículo 4. Principios generales [...] Responsabilidad
(vi) Principio de utilidad	Principio de orden original
(vii) Principio de acceso y circulación restringida	Ley 594 de 2000. Artículo 4. Principios generales [...] f) Administración y acceso. Información clasificada y reservada. Tablas de Control de Acceso. Índice de información clasificada y reservada
(viii) Principio de incorporación	Principio de procedencia

8 Al respecto confrontar: Ley 1581 de 2012; CORTE CONSTITUCIONAL. Sentencia T-729 de septiembre 05 de 2002. M.P. Eduardo Montealegre Lynett; CORTE CONSTITUCIONAL. Sentencia C-748 de octubre 06 de 2011. M.P. Jorge Ignacio Pretelt Chaljub. Fundamento jurídico 2.6.3. Consideraciones generales sobre el artículo 4.

(PRINCIPIOS) HABEAS DATA⁹ Ley 1581 de 2012. Artículo 4°. Principios para el tratamiento de datos personales. Corte Constitucional. Sentencia C-748 de 2011.	FUNCIÓN ARCHIVÍSTICA. Ley 594 de 2000. Artículo 4. Principios generales. Corte Constitucional. Sentencia T-656 de 2010.
(ix) Principio de temporalidad, caducidad ⁹ o derecho al olvido (Supresión de datos)	Retención documental. Valoración documental. Disposición final de los documentos, selección, eliminación, conservación permanente
(x) Principio de individualidad	Principio de procedencia
(xi) Principio de legalidad en materia de tratamiento de datos personales	Ley 594 de 2000. Artículo 4. Principios generales [...] Responsabilidad
(xii) Principio de transparencia o de libertad de acceso al titular	Ley 594 de 2000. Artículo 4. Principios generales [...] f) Administración y acceso
(xiii) Principio de seguridad	Ley 594 de 2000. Artículo 4. Principios generales [...] f) Administración y acceso
(xiv) Principio de confidencialidad	Ley 594 de 2000. Artículo 4. Principios generales [...] f) Administración y acceso. Información clasificada y reservada. Tablas de Control de Acceso. Índice de información clasificada y reservada

Fuente: Extracto del capítulo "Habeas Data y Función Archivística", de libro Documentos, mensajes de datos de Arturo Daniel López Caba.

En el anterior cuadro comparativo se identifican los principios del habeas data que guardan relación con la función archivística y cuáles no lo presentan, algunos de ellos se relacionan con artículos de la Ley 594 de 2000, Ley General de Archivos.

Del mismo, se puede concluir que, los principios del *Habeas Data*, se integran o complementan con principios archivísticos. Por tal motivo, es procedente dar aplicación a los principios del *Habeas Data* a los archivos, como lo ordena el artículo 15 de la Constitución Política de Colombia.

⁹ CORTE CONSTITUCIONAL. Sentencia C-185 de marzo 4 de 2003. M.P. Eduardo Montealegre Lynnet. Fundamento jurídico 9. "En este sentido la Corte le ha reconocido validez al principio de caducidad o de temporalidad de la información negativa, lo que implica que la información personal desfavorable al titular de la misma debe ser retirada de las bases de datos siguiendo criterios de razonabilidad y de oportunidad. Igualmente, ha considerado constitucionalmente inadmisibles la conservación indefinida de datos personales que revelen información negativa una vez hayan desaparecido las causas que justificaron su acopio y administración".

5.1.3. Para la Corte los principios del habeas data implica deberes constitucionales para las entidades que custodian y administran la información contenida en archivos y bases de datos. Así, dichas entidades deben observar una obligación general de seguridad y diligencia en la administración y conservación de los datos personales y una obligación específica de corregir e indemnizar los perjuicios causados por el mal manejo de la información¹⁰.

5.1.4. En este orden de ideas, debe resaltarse la importancia de que el acopio y la conservación de información se haga con sujeción a los principios del habeas data con el fin de garantizar su integridad y veracidad y así salvaguardar los demás derechos de los titulares de la información. Con frecuencia esta información es necesaria para acceder al goce efectivo de otros derechos fundamentales, toda vez que los datos personales, laborales, médicos, financieros y de otra índole que están contenidos en archivos y bases de datos, son la fuente de la información que se utiliza para evaluar el cumplimiento de los requisitos para el reconocimiento de derechos y prestaciones¹¹ (artículo 15).

En la jurisprudencia de la Corte Constitucional se encuentra la relación entre documentos (información) y datos personales, considerando que:

21. Los datos no corresponden al documento, sino que el documento contiene datos. Sin embargo, la producción exponencial de documentos que contienen datos socialmente relevantes demanda una debida custodia, organización, actualización y circulación¹².

3. Algunos supuestos en los que las entidades públicas deben anonimizar

La anonimización no debe ser una facultad discrecional de las entidades públicas, sino una obligación que se activa en supuestos específicos definidos por el ordenamiento jurídico vigente. A continuación, se describen los principales escenarios en los que su aplicación resulta procedente y, en varios casos, imperativa.

10 Sentencias T-718 de 2005 y C-1011 de 2008. Citada en: CORTE CONSTITUCIONAL. Sentencia T-926 de diciembre 06 de 2013. M.P. Mauricio González Cuervo.

11 CORTE CONSTITUCIONAL. Sentencia T-926 de diciembre 06 de 2013. M.P. Mauricio González Cuervo.

12 CORTE CONSTITUCIONAL. Sentencia T-216 de marzo 08 de 2004. M.P. Eduardo Montealegre Lynett. Fundamento jurídico 21.

Atención de solicitudes de acceso a la información pública

Cuando una entidad pública recibe una solicitud de acceso a la información, ya sea mediante derecho de petición, acción de tutela o solicitud formal en los términos de la Ley 1712 de 2014, y el documento requerido contiene datos personales no sujetos al régimen de transparencia, la anonimización constituye el mecanismo jurídicamente idóneo para dar respuesta sin vulnerar el derecho fundamental a la protección de datos personales. En estos casos, la entrega de una versión pública del documento, con los datos personales debidamente anonimizados, permite conciliar el principio de máxima publicidad con el deber de protección de la información personal.

Divulgación proactiva de información pública

Las entidades obligadas a publicar información de oficio en portales institucionales, plataformas de datos abiertos, informes de gestión o espacios de rendición de cuentas, deben garantizar que dicha publicación no incluya datos personales cuya divulgación resulte innecesaria para el cumplimiento del principio de transparencia. En estos supuestos, la anonimización opera como condición previa a la publicación, en estricta observancia del principio de finalidad y acceso y circulación restringida, consagrados en el artículo 4 de la Ley 1581 de 2012.

Mandato legal o reglamentario expreso

El ordenamiento jurídico colombiano contempla disposiciones específicas que ordenan la protección de determinadas categorías de información y de ciertos sujetos. Entre estos se encuentran las víctimas del conflicto armado, los menores de edad, los testigos y las fuentes en procesos judiciales, así como la información relacionada con actividades de inteligencia, contrainteligencia, seguridad y defensa nacional. En estos casos, la anonimización no es una opción sino una obligación legal directa, cuyo incumplimiento puede acarrear responsabilidad disciplinaria, administrativa y penal para los servidores públicos responsables.

Tratamiento con fines estadísticos, científicos o históricos

Conforme a lo dispuesto en el artículo 6 de la Ley 1581 de 2012, el tratamiento de datos personales sensibles es procedente cuando la finalidad es exclusivamente estadística, histórica o científica, siempre que se elimine toda posibilidad de identificación directa o indirecta del titular. La anonimización es, en estos supuestos, el requisito técnico que habilita jurídicamente dicho tratamiento y garantiza su licitud.

Procesos de desclasificación, verdad y memoria histórica

En el marco de la justicia transicional y los mecanismos de esclarecimiento de la verdad, la anonimización adquiere una dimensión jurídica y ética de especial relevancia. Aplica, entre otros, a archivos de inteligencia y contrainteligencia, expedientes de derechos humanos, fondos

documentales sobre el conflicto armado y acervos históricos que contienen información con impacto directo en derechos fundamentales.

El levantamiento de la reserva o la desclasificación de un documento no equivale automáticamente a la publicación íntegra del documento: antes de habilitar el acceso o consulta pública, la entidad debe identificar si el documento contiene información personal semiprivada, privada o sensible, cuya confidencialidad no está determinada por el período de reserva del documento, sino por la naturaleza del dato y los derechos del titular. La confidencialidad de los datos personales contenidos en un documento es de carácter indefinido, salvo que exista una habilitación legal expresa para su divulgación, que el titular haya otorgado su consentimiento, o sean aquellos que estén sujetos al principio de máxima publicidad. Por lo tanto, una vez levantada la reserva, o desclasificado el documento, se debe generar la identificación de la información personal que persiste protegida, y finalmente la anonimización como condición para la creación de la versión pública del documento.

Pruebas tecnológicas, desarrollo de software, uso de Inteligencia Artificial

En aplicación de los principios de privacidad por diseño y privacidad por defecto¹³, las entidades públicas que utilicen información institucional en ambientes de prueba, procesos de desarrollo de software, ejercicios de analítica interna, capacitaciones o pilotos tecnológicos, están obligadas a emplear datos anonimizados. El uso de datos personales reales en estos contextos carece de base legal y expone a la entidad a riesgos de tratamiento ilícito de datos.

Seguridad y defensa nacional

Aun cuando cierta información no revista carácter sensible de forma autónoma, el contexto operacional en el que se produce o utiliza puede conferirle un valor estratégico que justifica su anonimización. Este supuesto aplica cuando la divulgación de la información podría permitir la identificación de fuentes humanas, la correlación de operaciones, la exposición de funcionarios o la revelación de métodos, capacidades y medios empleados por los organismos de seguridad del Estado. En estos casos, la anonimización es una medida de seguridad operacional con fundamento en las normas que regulan la reserva en materia de inteligencia y defensa nacional.

13 Circular Externa 2 de 2024 de la Superintendencia de Industria y Comercio Lineamientos sobre el Tratamiento de Datos personales en Sistemas de Inteligencia Artificial.

Información clasificada y reservada

La Ley 1712 de 2014 establece dos categorías de información exceptuada del principio de máxima publicidad: la información clasificada, cuyo acceso puede ser rechazado cuando su divulgación cause un daño a derechos de personas naturales o jurídicas; y la información reservada, cuyo acceso puede ser denegado cuando su divulgación afecte intereses públicos expresamente protegidos por norma legal o constitucional. El artículo 18 de la citada ley protege el derecho a la intimidad, con las limitaciones propias de la condición de servidor público, el derecho a la vida, la salud o la seguridad, y los secretos comerciales, industriales y profesionales. En el segundo caso, el artículo 19 salvaguarda bienes jurídicos de interés general, como la defensa y seguridad nacional, la seguridad pública, las relaciones internacionales, la investigación y persecución de delitos y faltas disciplinarias, el debido proceso, la administración de justicia, los derechos de la infancia y la adolescencia, la estabilidad macroeconómica y financiera, y la salud pública.

Frente a estos supuestos, la anonimización se erige como el mecanismo técnico y jurídico que permite a las entidades públicas entregar una versión pública del documento sin incurrir en las causales de daño previstas por la ley. En lugar de denegar íntegramente el acceso, la entidad puede suprimir o proteger los elementos que activan la excepción, garantizando así el acceso a la información no exceptuada y dando cabal cumplimiento al principio de máxima publicidad consagrado en el artículo 2 de la misma ley. Esta aproximación es concordante con lo dispuesto por el Decreto 1081 de 2015 y con la jurisprudencia de la Corte Constitucional y el Consejo de Estado en la materia. Cabe destacar que, conforme al párrafo del artículo 18, las excepciones por daño a derechos de personas naturales o jurídicas tienen duración ilimitada, salvo que el titular haya consentido en la revelación de sus datos o que la información haga parte del régimen de publicidad aplicable. En consecuencia, la obligación de anonimizar en estos supuestos no está sujeta a plazos de desclasificación y debe observarse con independencia de la antigüedad del documento.

Los supuestos descritos evidencian que la anonimización no responde a un criterio único ni a una lógica uniforme, sino que se activa en contextos normativos, institucionales y operacionales diversos, cada uno con sus propias exigencias técnicas y jurídicas. La anonimización es, en cada uno de estos escenarios, la respuesta técnicamente viable y jurídicamente válida a esa tensión.

4. Principios aplicables al proceso de anonimización

La aplicación de técnicas de anonimización debe realizarse en coherencia de los principios constitucionales y legales orientados a la protección de la dignidad humana, en especial los principios que rigen el tratamiento de datos personales, el acceso a la información pública, la gestión documental y archivística, y la protección y promoción de los derechos humanos. Este marco se complementa con los estándares internacionales establecidos en los Principios de Johannesburgo sobre Seguridad Nacional, Libertad de Expresión y Acceso a la Información (1996) y los Principios de Tshwane sobre Seguridad Nacional y el Derecho a la Información (2013), que consolidan la regla de máximo acceso con mínima restricción, y reconocen la protección de la identidad de víctimas, testigos e informantes como condición para hacer efectivo ese acceso. Estos principios, en conjunto, no son opcionales ni meramente orientadores: constituyen el marco jurídico y ético vinculante dentro del cual deben inscribirse todas las decisiones técnicas del proceso de anonimización.

De manera complementaria, y sin que ello implique la adopción de un marco jurídico distinto al colombiano, se incorporan como criterios técnicos de buenas prácticas los principios desarrollados por la Agencia Española de Protección de Datos en su Guía de Anonimización¹⁴. Estos principios no tienen carácter obligatorio ni forman parte del ordenamiento jurídico colombiano; sin embargo, constituyen referentes técnicos reconocidos internacionalmente, que permiten orientar la aplicación de técnicas de anonimización de forma rigurosa y sistemática. Su adopción debe entenderse de manera armónica con las disposiciones de la normativa colombiana vigente, particularmente en lo relativo a la protección de los derechos de los titulares y la prevención de riesgos de reidentificación. En ese marco, los principios orientadores del proceso de anonimización son los siguientes:

Principio de acción proactiva. La protección de la privacidad constituye el objetivo central del proceso de anonimización y su gestión debe ser anticipatoria en lugar de reactiva. Las medidas necesarias para garantizar la privacidad de las personas deben adoptarse desde el inicio del diseño del sistema de información o del proceso de anonimización, y no como respuesta correctiva

¹⁴ El principio de privacidad por diseño y por defecto ha sido desarrollado por la Superintendencia de Industria y Comercio como Autoridad de Protección de Datos Personales, a través de sus diferentes guías en materia de protección de datos personales. También, a través de la Circular Externa 002 de 2024 sobre protección de datos en sistemas de inteligencia artificial.

a fallos o daños posteriores. En consecuencia, es esencial identificar y eliminar, desde el origen, toda posible vía de reidentificación de los titulares en los datos anonimizados.

Principio de privacidad por defecto. El diseño de cualquier sistema de información o proceso de anonimización debe tener como punto de partida la protección de la confidencialidad de los titulares. Ello implica definir desde el inicio el nivel de detalle que deben tener los datos resultantes, garantizando que la configuración predeterminada del sistema sea siempre la más protectora de la privacidad, sin requerir intervención adicional por parte del usuario o del responsable del tratamiento.

Principio de privacidad objetiva. El responsable del tratamiento debe asumir y gestionar el riesgo residual de reidentificación como un elemento inherente al proceso de anonimización. Dicho riesgo debe ser evaluado, documentado y comunicado al destinatario de la información anonimizada. Cuando los datos anonimizados sean de acceso público, el nivel de riesgo residual deberá comunicarse también a las personas o entidades que hagan uso de esa información, permitiendo decisiones informadas sobre su utilización.

Principio de plena funcionalidad. Desde el diseño del proceso de anonimización se debe garantizar que los datos resultantes conserven su utilidad para los fines que motivaron su tratamiento, sin que las técnicas aplicadas introduzcan distorsiones que comprometan la validez del análisis o la integridad de la información. La anonimización no debe sacrificar la funcionalidad de los datos más allá de lo estrictamente necesario para proteger la identidad de los titulares.

Principio de privacidad en el ciclo de vida de la información. Las medidas de protección de la privacidad deben aplicarse de manera continua durante todo el ciclo de vida de la información, desde su producción o recepción en su forma original, a lo largo de todas las etapas de tratamiento, hasta su disposición final. La anonimización no es un evento puntual sino un atributo que debe mantenerse y verificarse en cada fase del proceso documental.

Principio de información y formación. El diseño del sistema de información o del proceso de anonimización debe contemplar desde su concepción las necesidades de capacitación y comprensión de todos los actores involucrados. Cada perfil (técnico, jurídico, archivístico o directivo) requiere un nivel de conocimiento específico sobre el proceso, sus implicaciones y sus responsabilidades, lo cual debe ser previsto como parte integral del diseño institucional del proceso de anonimización.

5. Marco legal de la anonimización en Colombia

El marco legal de la anonimización en Colombia viene dado principalmente por el régimen normativo de protección de datos personales, en especial la Ley 1581 de 2012 y sus decretos reglamentarios. Esta define los principios para el tratamiento de datos personales tanto en archivos como en bases de datos, los tipos de datos personales, los roles que interactúan en el tratamiento de los datos, y las obligaciones de los administradores de esas bases de datos, así como los derechos de los titulares de la información.

De manera paralela, se encuentra el régimen normativo de transparencia y acceso a la información pública, esto es, la Ley 1712 de 2014 y sus decretos reglamentarios. Este régimen establece los principios para el acceso a la información pública, define las reglas para su ejercicio, y determina las condiciones en que opera la información pública clasificada y reservada. Su decreto reglamentario, el Decreto 103 de 2015, compilado en el Decreto 1081 de 2015, establece en su artículo 31 las condiciones en que opera la anonimización de la información.

5.1 Ley 1581 de 2012. Régimen general de la protección de datos personales en Colombia

La Ley 1581 de 2012, que regula la protección de datos personales en Colombia, contiene varios principios y disposiciones que sustentan directamente la necesidad de aplicar procesos de anonimización.

En primer lugar, el literal f) del artículo 4 establece que los datos personales, salvo aquellos de naturaleza pública, no pueden estar disponibles en internet, ni en otros medios de divulgación o comunicación masiva, a menos que el acceso esté técnicamente controlado y limitado exclusivamente a los titulares o a terceros debidamente autorizados por la ley. Este mandato implica que, cuando la información deba hacerse pública, se eliminen o transformen los datos que permitan identificar a una persona.

El literal g) del artículo 4 consagra el principio de seguridad, el cual obliga a los responsables y encargados del tratamiento implementar medidas técnicas, humanas y administrativas adecuadas para proteger la información personal frente a riesgos de adulteración,

pérdida, consulta, uso o acceso no autorizado o fraudulento. La anonimización se convierte así en una medida técnica idónea para mitigar estos riesgos.

Por su parte, el literal h) del artículo 4 desarrolla el principio de confidencialidad, imponiendo a todas las personas que intervienen en el tratamiento de datos personales el deber de garantizar la reserva de la información, incluso después de finalizada su relación con las actividades que dieron lugar al tratamiento. Este deber se refuerza mediante técnicas que impidan la reidentificación de los titulares.

Merece especial atención, el numeral 2 del artículo 6, debido a que dispone que el tratamiento de datos sensibles está prohibido, salvo cuando tenga una finalidad histórica, estadística o científica. En estos eventos, la ley exige adoptar medidas conducentes a la supresión de la identidad de los titulares. Esto constituye un mandato expreso de anonimización.

5.2 Ley 1712 de 2014. Régimen general de transparencia y acceso a la información pública

En materia de acceso a la información pública, la Ley 1712 de 2014 establece límites claros a la divulgación cuando están en juego derechos fundamentales. El artículo 18 señala que la información pública clasificada se encuentra exceptuada del acceso al público, cuando su divulgación pueda afectar el derecho a la intimidad de las personas, incluidos los servidores públicos en los límites propios de su función, así como los derechos a la vida, la salud o la seguridad, y los secretos comerciales, industriales y profesionales. Estas excepciones tienen carácter indefinido y no resultan aplicables únicamente cuando el titular consiente expresamente la divulgación o cuando los datos fueron entregados en el marco de información, sometida al régimen de publicidad.

En desarrollo de esta ley, deben considerarse también los supuestos de restricción relacionados con: el derecho a la intimidad; la vida, la salud y la seguridad; los secretos comerciales, industriales y profesionales; los datos de niños, niñas y adolescentes; la defensa y seguridad nacional; la seguridad pública; las relaciones internacionales; la prevención, investigación y persecución de delitos y faltas disciplinarias, mientras no se hagan efectivas las medidas correspondientes; el debido proceso y la igualdad de las partes; la administración efectiva de justicia; los derechos de la infancia y la adolescencia; la estabilidad macroeconómica y financiera del país, y la salud pública.

El Decreto 103 de 2015, compilado en el Decreto 1081 de 2015 (artículo 2.1.1.1.1.) y en el Decreto 1080 de 2015 (artículo 2.8.4.1.1.), establece que permitir el acceso a un dato semiprivado, privado o sensible no le quita su carácter de información clasificada ni implica su desprotección. El acceso a este tipo de información solo es posible por decisión de una autoridad jurisdiccional o de una autoridad pública o administrativa competente en ejercicio de sus funciones.

El artículo 31 del Decreto 103 de 2015, compilado en el Decreto 1081 de 2015 (artículo 2.1.1.1.1.) y en el Decreto 1080 de 2015 (artículo 2.8.4.3.2.), introduce un criterio operativo clave para la gestión de documentos con información mixta: cuando un mismo acto o documento contenga información divulgable e información clasificada o reservada, el sujeto obligado debe revelar los datos no protegidos y justificar, con fundamento constitucional y legal, la reserva o clasificación de la información que no puede divulgarse. Para ello, la norma autoriza expresamente el uso de mecanismos como el tachado, la anonimización, la transliteración o la edición del documento, así como la apertura de expedientes separados u otras medidas que permitan garantizar el acceso a la información no protegida, teniendo en cuenta el formato y el medio de conservación del documento.

5.3 Legislación complementaria

El marco legal de la anonimización no puede ser analizado sin considerar el régimen general de archivos en Colombia, en especial la Ley 594 de 2000, la cual establece que:

Todas las personas tienen derecho a consultar los documentos de archivos públicos y a que se les expida copia de los mismos, siempre que dichos documentos no tengan carácter reservado conforme a la Constitución o a la ley. Las autoridades responsables de los archivos públicos y privados garantizarán el derecho a la intimidad personal y familiar, honra y buen nombre de las personas y demás derechos consagrados en la Constitución y las leyes (artículo 27).

Este mandato no es exclusivo del régimen general de archivos. El deber de proteger la intimidad y los datos personales contenidos en documentos públicos se extiende a múltiples sectores del Estado, cada uno con sus propias disposiciones específicas, tal como se ilustra a continuación.

Sector / Norma	Relación con la anonimización
Normativa de Archivo y Gestión Documental (Integridad del Patrimonio)	• Ley 594 de 2000: Ley General de Archivos • Acuerdo 001 de 2024 del AGN: Reglamento General de Archivos que establece las condiciones de acceso y consulta, protegiendo el patrimonio documental • Decreto 1400 de 2025: Ordena explícitamente mantener los originales inalterados durante el proceso de desclasificación del DAS • Decreto 1080 de 2015, Decreto Único del Sector Cultura, artículos 2.8.2.1.1. y siguientes: regula el patrimonio archivístico
Inteligencia y contrainteligencia	• Ley 1621 de 2013 (Art. 33): Establece la reserva legal por 30 años y la protección de la identidad de los agentes, los medios, los métodos y las fuentes • Decreto 1070 de 2015 y Decreto 857 de 2014: Regulan la clasificación de documentos de inteligencia y ordenan proteger la identidad funcional de los agentes • Decreto 2149 de 2017: Crea el Sistema Nacional de Depuración de Datos y Archivos de Inteligencia y Contrainteligencia, y obliga a proteger datos que pongan en riesgo a fuentes, agentes, medios, métodos, procesos y procedimientos de inteligencia y contrainteligencia
Derechos de niños, niñas y adolescentes (protección reforzada)	• Ley 1098 de 2006, Código de la Infancia y la Adolescencia, en especial su artículo 47: Prohíbe divulgar datos que identifiquen a menores víctimas, autores o testigos de hechos delictivos, imponiendo una obligación reforzada de anonimización en cualquier documento que los involucre
Justicia transicional y derechos de las víctimas (verdad vs. privacidad)	• Ley 1448 de 2011 (Ley de Víctimas): Constituye la base normativa para la protección de la información contenida en el Registro Único de Víctimas • Jurisprudencia de la JEP (Autos 001, 073, 182, 778): Ordenan medidas cautelares sobre archivos del DAS y definen protocolos de acceso para el Sistema Integral de Verdad • Principios de Joinet y Tshwane: Estándares internacionales de referencia en materia de lucha contra la impunidad y acceso a archivos de represión, ampliamente citados en los instrumentos que regulan los procesos de justicia transicional en Colombia

Sector / Norma	Relación con la anonimización
	• Archivo General de la Nación y CNMH (julio de 2022). Protocolo de Gestión Documental de los Archivos Referidos a las Graves y Manifiestas Violaciones a los Derechos Humanos e Infracciones al Derecho Internacional Humanitario, Ocurredas con Ocasión del Conflicto Armado Interno, Versión II
Normativa sectorial específica (salud y estadística)	• Ley 2335 de 2023 (Reserva Estadística): “Por la cual se expiden disposiciones sobre las estadísticas oficiales en el país” • Ley 23 de 1981 (artículo 33 y siguientes Reserva de la Historia Clínica”
Disposiciones relativas a la clasificación y reserva de la información	Los artículos 18 y 19 de la Ley 1712 de 2014 regulan las categorías de información clasificada y reservada. El artículo 18 protege derechos de personas naturales o jurídicas, como la intimidad, la vida, la salud, la seguridad y los secretos empresariales, estos últimos regulados principalmente por la Decisión 486 de la Comunidad Andina y por el artículo 16 de la Ley 256 de 1996. El artículo 19 establece excepciones por daño a intereses públicos esenciales. Estos supuestos fueron desarrollados en detalle en la sección 4 del presente documento, a la cual se remite para su análisis completo
Estándares internacionales de referencia	• Principios de Johannesburgo (1996) y Principios de Tshwane (2013): establecen la regla de máximo acceso con mínima restricción y reconocen la protección de identidad de víctimas, testigos e informantes como condición para hacer efectivo ese acceso • Principios Orentlicher (ONU, 2005): fundamentan el deber estatal de preservar archivos como garantía del derecho a la verdad • Directrices OCDE (1980/2013) y estándares ONU (Resolución 68/167 de 2013): establecen el principio de minimización del dato como fundamento internacional de la anonimización

Fuente: elaboración propia.

6. La información no estructurada en los documentos de archivo

En el ámbito archivístico y de gestión documental, es necesario distinguir entre dato, formato y soporte. El dato corresponde a la unidad mínima de información; el formato al modo en que dicha información se organiza y presenta; y el soporte al medio físico o digital en el cual se contiene (papel, archivo electrónico, grabación sonora, registro audiovisual, entre otros). Esta distinción es relevante en materia de anonimización, dado que **el riesgo de identificación no depende únicamente del contenido informativo del documento, sino también del formato en que dicha información se organiza y del soporte en que se conserva, factores que condicionan tanto las técnicas de anonimización aplicables como el nivel de intervención requerido.**

De acuerdo con la Guía de Anonimización de Datos Estructurados-AGN (2020), los datos no estructurados son aquellos cuya organización y presentación no está guiada por un modelo o esquema predefinido. En esta categoría se incluyen, entre otros, textos narrativos, imágenes, audios, videos, contenidos publicados en redes sociales, así como documentos en formatos como PDF o Word, cuando no responden a una estructura tabular o de campos previamente definidos.

Uno de los principales retos de la información no estructurada radica en la diversidad de elementos que confluyen en un mismo documento: sujetos con distintas condiciones jurídicas, temáticas heterogéneas, etapas variadas de la actuación administrativa y contextos fácticos específicos. Esta heterogeneidad dificulta la estandarización de reglas automáticas de anonimización, especialmente cuando la identificación puede derivarse no solo de datos directos, sino de la narrativa, la combinación de atributos o el contexto territorial e institucional.

En el contexto de los archivos públicos, el expediente constituye la unidad documental por excelencia en la que coexisten múltiples formatos no estructurados: textos narrativos, imágenes, registros audiovisuales, documentos digitalizados y anotaciones manuscritas, entre otros. Esta coexistencia incrementa significativamente el riesgo de identificación por combinación de elementos, aun cuando cada componente considerado de forma aislada no permita identificar al titular. Por ello, la anonimización en el contexto archivístico debe abordarse desde una perspectiva de expediente y no solo de documento individual.

En los formatos no estructurados, la información personal se encuentra integrada al contenido y no puede aislarse mediante mecanismos automáticos simples. Por ello, aun cuando se empleen herramientas tecnológicas de detección o supresión automática, deberá realizarse

una revisión humana y contextual por parte de un equipo que conozca la naturaleza de la información tratada, el marco normativo aplicable y los riesgos de reidentificación asociados. La automatización puede apoyar el proceso, pero no sustituye el análisis técnico y jurídico necesario para garantizar una anonimización efectiva.

6.1 Formatos no estructurados que pueden contener datos personales o información sensible

A continuación, se identifican los principales formatos no estructurados presentes en los acervos documentales de las entidades públicas, junto con los elementos de cada uno que pueden contener datos personales o información sensible susceptible de anonimización.

Documentos físicos (no estructurados en origen). Soportes en papel cuya información requiere tratamiento manual antes de su reproducción o digitalización. En esta categoría se encuentran los expedientes físicos, las actas manuscritas, los informes de misión o visitas de campo, las anotaciones, los formularios diligenciados a mano y los soportes anexos que integran el expediente.

Documentos de texto libre (digitales). Archivos en los que la información personal se encuentra dispersa en narrativas, sin una estructura fija de campos. Incluyen documentos en formato Word y/o PDF, informes, memorandos, oficios y actas; correos electrónicos institucionales con sus hilos de conversación y archivos adjuntos; informes narrativos de carácter técnico, misional, disciplinario o de investigación, y expedientes electrónicos cuyo contenido principal es texto libre.

Formatos multimedia (audio y video). Soportes audiovisuales en los que la información personal se encuentra embebida en señales acústicas, imágenes en movimiento y contexto. Comprenden grabaciones de audio, llamadas, entrevistas, audios de atención al ciudadano, registros operativos, en los que la voz humana es susceptible de identificación, mediante análisis acústico y el sonido ambiente puede contener menciones de nombres, ubicaciones o rutinas. También incluyen grabaciones de video de registros operativos, audiencias e institucionales, así como imágenes de sistemas de videovigilancia (CCTV) que registran personas, comportamientos, desplazamientos y rasgos físicos.

Imágenes y formatos visuales. Soportes gráficos en los que los datos personales se encuentran integrados al contenido visual. Se incluyen fotografías fijas de personas, actuaciones administrativas o evidencias; escaneos de documentos de identidad, carnés, fichas técnicas o

expedientes físicos digitalizados; e imágenes que contengan datos biométricos como huellas dactilares, iris, o señales físicas visibles como tatuajes, cicatrices o marcas de nacimiento.

Registros digitales incorporados a expedientes. Información recolectada desde plataformas externas o entornos abiertos, que ha sido incorporada formalmente a expedientes o sistemas institucionales. Comprende publicaciones en redes sociales, chats y comunicaciones de mensajería instantánea incorporadas como evidencia, y capturas de pantalla de aplicaciones, perfiles o interacciones digitales.

6.2 Datos específicos dentro de formatos no estructurados

Los formatos no estructurados pueden contener distintas categorías de datos personales, que permiten la identificación directa o indirecta de personas. A continuación, se describen las principales:

Datos de localización y contexto espacial. Direcciones físicas, referencias territoriales, coordenadas GPS o fondos reconocibles en imágenes y videos. Implican riesgos de rastreo y localización de personas.

Identificadores de movilidad y transporte. Placas o matrículas de vehículos visibles en fotografías, videos o grabaciones de CCTV. Facilitan la identificación indirecta y el seguimiento de desplazamientos.

Identificadores documentales. Firmas manuscritas, iniciales o rúbricas presentes en documentos físicos o digitalizados, que permiten la atribución de autoría e identidad.

Identificadores tecnológicos. Direcciones IP, URL o enlaces registrados en informes técnicos, bitácoras o correos electrónicos. Permiten la trazabilidad de actividades y la correlación de información entre sistemas.

Identificadores biométricos. El rostro y la voz constituyen los identificadores de mayor riesgo en los formatos no estructurados, por cuanto permiten la identificación unívoca de las personas y son técnicamente difíciles de disociar del contenido del documento. Por esta razón, las fotografías, videos, grabaciones de CCTV y registros de voz deben ser tratados como formatos no estructurados de alto riesgo, que requieren técnicas especializadas de anonimización, aplicadas de manera consistente durante todo su ciclo de uso y reproducción.

6.3 Información no estructurada que se utiliza en las entidades públicas

Las entidades públicas producen, reciben y conservan información no estructurada en el marco de sus funciones misionales, administrativas y de gestión. A continuación, se presentan, a manera de ejemplo, las principales tipologías documentales en las que este tipo de información suele estar presente:

Documentos de gestión de personal y función pública. Expedientes de personal que incluyen hojas de vida, soportes de vinculación, evaluaciones, datos de contacto personal, información familiar. También se incluyen actos de vinculación, retiro o situación administrativa, cuyas resoluciones y documentos soporte, pueden contener información médica o datos reservados, así como actuaciones disciplinarias internas con documentos narrativos que incorporan testimonios, descargos, pruebas y datos de terceros.

Documentación asociada a la contratación estatal. Documentos generados en el marco de la relación contractual con personas naturales o jurídicas, que incorporan datos personales y financieros. Comprenden contratos con personas naturales que contienen identificadores personales, información de localización y datos económicos; soportes de ejecución contractual como informes, cuentas de cobro y anexos que pueden incluir datos de seguridad social e ingresos; y documentos financieros de soporte como certificaciones bancarias, pagos y comprobantes asociados a la ejecución del contrato.

Documentos relacionados con salud, bienestar y protección social. Actuaciones que revelan información sobre el estado físico, mental o condiciones de vulnerabilidad de las personas. Se incluyen expedientes y conceptos médicos como historias clínicas, valoraciones, diagnósticos o conceptos técnicos en salud, así como actos de reconocimiento de prestaciones relacionados con pensiones, invalidez, indemnizaciones o ayudas humanitarias, que permiten inferir condiciones de salud o discapacidad.

Documentos de seguridad y defensa nacional, inteligencia y contrainteligencia. Información de alta sensibilidad producida por entidades del sector defensa o seguridad del Estado. Comprende informes de inteligencia y contrainteligencia con documentos narrativos que contienen perfiles, alias, descripciones físicas, ubicaciones y patrones de comportamiento, así como soportes de gastos reservados con documentación contable y administrativa asociada a fuentes humanas o actividades operativas, cuya divulgación implica riesgos graves para la seguridad del Estado y de las personas involucradas.

Documentos que involucran poblaciones de especial protección. Actuaciones administrativas, judiciales o misionales que involucran sujetos con protección constitucional reforzada. Se incluyen documentos relacionados con niños, niñas y adolescentes en actuaciones administrativas o judiciales que deben suprimir cualquier dato identificador, así como registros y expedientes de víctimas que contienen relatos de hechos victimizantes, información sensible y contextos de riesgo de revictimización. Igualmente, se incluyen los documentos que involucran a pueblos y comunidades étnicas indígenas, afrocolombianas, raizales, palenqueras y Rom.

Actuaciones administrativas generales y de gestión institucional. Documentos producidos en la operación cotidiana de las entidades, usualmente en formato narrativo. Comprenden actos administrativos que contengan datos personales; informes de gestión, control y auditoría con reportes narrativos sobre resultados, hallazgos o evaluaciones; actas de reuniones y comités con registros de deliberaciones, decisiones y, en algunos casos, información reservada; y peticiones, quejas, reclamos y solicitudes ciudadanas recibidas en texto libre.

Actuaciones judiciales y disciplinarias. Expedientes y actuaciones tramitadas ante autoridades judiciales o disciplinarias. Incluyen expedientes procesales con denuncias, pruebas, testimonios, interrogatorios y providencias; procesos disciplinarios y sancionatorios con sus fases probatorias y decisorias; y testimonios y versiones libres en forma de relatos extensos, escritos o grabados, de víctimas, testigos o comparecientes.

Archivos del conflicto armado, derechos humanos y memoria histórica. Constituyen uno de los conjuntos documentales de mayor sensibilidad en el contexto colombiano, dado que su contenido involucra directamente derechos fundamentales de las víctimas y el deber del Estado de garantizar verdad, justicia y reparación. En esta categoría se encuentran relatos de víctimas en forma de declaraciones, cartas y solicitudes; informes sobre violaciones a los derechos humanos; registros de desaparición forzada, tortura y violencia sexual; expedientes de restitución de derechos o tierras, y archivos notariales con sus respectivos anexos documentales.

Los ejemplos anteriores ilustran tipologías de información no estructurada presentes en los acervos de las entidades públicas. Su inclusión en este listado no implica que dicha información deba ser anonimizada de manera automática, sino que corresponde a casos en los que es posible identificar la presencia de datos personales que podrían requerir tratamiento. En este contexto, se recomienda que las entidades públicas realicen diagnósticos institucionales que les permitan identificar, de manera proactiva, los tipos de información que deben ser objeto de anonimización. Si bien este proceso constituye una excepción y no la regla general, su aplicación inadecuada puede afectar la calidad de la información y, en algunos casos, el acceso efectivo a la

misma. Para orientar ese diagnóstico, los presentes lineamientos establecen los criterios, las técnicas y los procedimientos que deben seguirse cuando la anonimización resulte procedente.

7. Tipos de datos personales sujetos a anonimización, identificadores directos e indirectos

Una vez identificados los formatos y tipologías documentales en los que puede presentarse información no estructurada, resulta necesario precisar qué tipos de datos personales pueden estar contenidos en ellos y cuál es su nivel de protección jurídica, a efectos de determinar el alcance de la anonimización requerida. Esta sección complementa la perspectiva técnica y archivística de la sección anterior con una perspectiva jurídica, que permite establecer qué datos deben ser objeto de tratamiento prioritario y con qué nivel de protección.

Según la Guía de Anonimización de Datos Estructurados del AGN, los identificadores directos son características o atributos distintivos que por sí mismos permiten la identificación de una persona natural o jurídica de manera inequívoca, como el nombre o el número de cédula. Los identificadores indirectos, también denominados cuasi-identificadores, son atributos que por sí solos no identifican a una persona, pero que relacionados o en combinación con otros identificadores indirectos podrían permitir su identificación dentro de un conjunto de datos, como ocurre con la fecha de nacimiento combinada con el código postal.

Es importante advertir que la clasificación contenida en la tabla siguiente no debe interpretarse de manera literal ni estática. Algunos datos personales son dinámicos en cuanto a su clasificación: dependiendo del contexto, la finalidad del tratamiento y la combinación con otros datos, un dato que en principio es público puede adquirir el carácter de semiprivado, privado o incluso sensible. Este es el caso, por ejemplo, de los nombres y apellidos, las imágenes, la nacionalidad o la edad, cuya clasificación puede variar según el contexto específico en que se encuentren.

Clasificación	Tipo de datos personales	Fuente
PÚBLICO	Nombres y apellidos	Artículo 10, Ley 1581 de 2012
PÚBLICO	Tipo de identificación, número de identificación	Artículo 51, Código Electoral
PÚBLICO	Fecha y lugar de expedición	Artículo 51, Código Electoral
PÚBLICO	Estado civil	Artículo 10, Ley 1581 de 2012
PÚBLICO	Sexo	Artículo 10, Ley 1581 de 2012
PÚBLICO	Nacionalidad	Artículo 10, Ley 1581 de 2012
PÚBLICO	Correo electrónico institucional-corporativo	Criterio propio
PÚBLICO	Teléfono y celular corporativo	Resolución 24109 de 2016
PÚBLICO	Dirección de trabajo	Artículo 3, Ley 1581 de 2012
PÚBLICO	Datos relativos a la actividad comercial	Artículo 3, Ley 1581 de 2012
PÚBLICO	Datos relativos a la actividad profesional	Artículo 3, Ley 1581 de 2012
PÚBLICO	Datos generales relacionados con afiliación al sistema integral de seguridad social en salud (EPS)	Administradora de los recursos del sistema general de seguridad social en salud
SEMIPRIVADO	Lugar y fecha de nacimiento	Artículo 51, Código Electoral
SEMIPRIVADO	Edad	Artículo 51, Código Electoral
SEMIPRIVADO	Correo electrónico personal	Criterio propio
SEMIPRIVADO	Celular y teléfono personal	Criterio propio
SEMIPRIVADO	Datos financieros, crediticios y/o derechos de carácter económico y datos patrimoniales	Artículo 3, Ley 1266 de 2008
SEMIPRIVADO	Datos socioeconómicos: estrato, propiedad de vivienda, bienes muebles e inmuebles, ingresos, egresos, inversiones	Criterio propio
SEMIPRIVADO	Datos de información tributaria	Artículo 583, Estatuto Tributario

Clasificación	Tipo de datos personales	Fuente
SEMIPRIVADO	Datos relacionados con la actividad económica de la persona	Artículo 583, Estatuto Tributario
SEMIPRIVADO	Datos relacionados con la historia laboral: experiencia, cargo, fechas de ingreso y retiro, anotaciones, llamados de atención	Artículo 24, Ley 1437 de 2011
SEMIPRIVADO	Datos relacionados con el nivel educativo, capacitación e historial académico	Concepto 14-268304 de la SIC
SEMIPRIVADO	Datos de antecedentes judiciales y/o disciplinarios	Sentencia T-020 de 2014
PRIVADO	Datos personales de acceso a sistemas de información: usuarios, IP, claves, perfiles	Criterio propio
PRIVADO / SEMIPRIVADO / SENSIBLE	Imagen (fotografía y video). Su clasificación depende del contexto y del tipo de imagen: una fotografía institucional puede ser pública, mientras que una imagen que revele condiciones de salud, origen étnico o situación de vulnerabilidad tendrá carácter sensible. Se remite a la advertencia sobre la dinamicidad de los datos contenida en la introducción de esta sección	Concepto SIC número 7075 de 2018
SENSIBLE	Datos biométricos: huella, ADN, voz, cuando permitan identificar a una persona ante un sistema de información	Artículo 3, Ley 1581 de 2012

Clasificación	Tipo de datos personales	Fuente
SENSIBLE	Datos de descripción morfológica: color de piel, iris, cabello, señales particulares, estatura, peso, complexión	Artículo 3, Ley 1581 de 2012
SENSIBLE	Datos relacionados con la salud de la persona	Artículo 3, Ley 1581 de 2012
SENSIBLE	Datos sobre estado de salud que incluyan resultados de pruebas, laboratorios, diagnósticos médicos, psicológicos, psiquiátricos, medicamentos y tratamientos terapéuticos	Artículo 3, Ley 1581 de 2012
SENSIBLE	Datos relacionados con pertenencia a sindicatos, organizaciones sociales, de derechos humanos, religiosas o políticas; convicciones religiosas, filosóficas o políticas	Artículo 3, Ley 1581 de 2012 Artículo 1, Ley 13 de 1972
SENSIBLE	Datos de preferencia, identidad y orientación sexual; origen étnico-racial	Artículo 3, Ley 1581 de 2012
SENSIBLE	Población en condición de vulnerabilidad	Ley 1448 de 2011
SENSIBLE	Datos sobre personas en situación de discapacidad	Ley 1618 de 2013

Fuente: elaboración propia.

La tabla de clasificación jurídica anterior debe leerse de manera articulada con la tabla de identificadores directos e indirectos contenida en la Guía de Anonimización de Datos Estructurados-AGN (2020), disponible en el repositorio institucional del Archivo General de la Nación. Mientras la tabla de esta sección clasifica los datos personales según su nivel de protección jurídica, la tabla de la guía permite identificar operativamente qué atributos, considerados de forma individual o en combinación, pueden permitir la identificación de una

persona y deben ser objeto de tratamiento en el proceso de anonimización. Ambas herramientas son complementarias y deben utilizarse conjuntamente al momento de diseñar y aplicar las técnicas de anonimización.

7.1 Categorías de datos personales según su nivel de protección

El régimen normativo de protección de datos personales en Colombia, y la jurisprudencia de la Corte Constitucional, establece una clasificación de los datos personales según el nivel de protección que requieren y las condiciones bajo las cuales pueden ser recolectados y tratados. Comprender esta clasificación es esencial para determinar el alcance de la anonimización requerida en cada caso.

Los datos públicos son aquellos que requieren circular libremente para facilitar el contacto con la sociedad. Su tratamiento no requiere autorización del titular, pero este puede oponerse a su tratamiento, salvo que no sea obligatorio por ley dicho tratamiento. El carácter público de un dato no es absoluto: dependiendo del contexto en que se encuentre y de su combinación con otros datos, un dato originalmente público puede derivar en una afectación a la privacidad del titular que justifique su anonimización.

Los datos semiprivados son aquellos que no tienen naturaleza íntima ni reservada, pero tampoco son de libre circulación. Su conocimiento o divulgación puede interesar no solo al titular, sino a cierto sector o grupo de personas o a la sociedad en general. A diferencia de los datos públicos, su tratamiento requiere autorización del titular, aunque pueden ser solicitados de manera obligatoria cuando existe una base legal que lo justifique.

Los datos privados son aquellos que por su naturaleza íntima o reservada solo son relevantes para el titular.

Los datos sensibles son aquellos que afectan la intimidad del titular o cuyo uso indebido puede generar discriminación. Gozan del mayor nivel de protección en el ordenamiento jurídico colombiano y su tratamiento requiere autorización expresa del titular. No pueden ser solicitados de manera obligatoria, salvo que una norma legal expresamente establezca esa obligación. En tales casos, la autorización del titular sigue siendo necesaria.

Finalmente, **los datos de niños, niñas y adolescentes** merecen una mención especial por la protección constitucional reforzada que les otorga el ordenamiento jurídico. Los datos semiprivados y públicos de menores de edad pueden solicitarse de manera obligatoria, pero

siempre requieren autorización de su representante legal. Los datos privados y sensibles de niños, niñas y adolescentes no pueden solicitarse de manera obligatoria, salvo que la ley expresamente lo exija, caso en el cual la autorización del representante legal es igualmente indispensable. En el contexto archivístico, cualquier documento que involucre a un menor de edad debe ser tratado con el más alto estándar de protección, independientemente de la clasificación individual de cada dato que contenga.

8. Sujetos, espacios y contextos objeto de protección mediante procesos de anonimización

Los procesos de anonimización deben considerar de manera integral los sujetos involucrados, los espacios en los que se desarrollan los hechos y las circunstancias que pueden facilitar la identificación directa o indirecta de personas naturales. La determinación de quién debe ser protegido no depende únicamente de la naturaleza pública o privada del documento, sino del análisis contextual del riesgo que la divulgación puede generar sobre la vida, la integridad, la intimidad, la dignidad o la seguridad de los titulares

8.1 Categoría de sujetos protegidos

Los sujetos de protección en los procesos de anonimización comprenden a cualquier persona natural, cuya información personal no esté sometida al régimen de transparencia y publicidad, razón por la cual no puede ser divulgada sin una habilitación legal expresa.

Cuando la información personal se encuentre incorporada dentro de documentos, archivos o registros que tengan carácter público, la entidad deberá definir criterios claros de anonimización orientados a suprimir, disociar o alterar la identidad del titular y cualquier elemento que permita su identificación directa o indirecta.

La anonimización debe aplicarse tanto a quienes sufren un daño o se encuentran en situación de especial vulnerabilidad, como víctimas, niños, niñas y adolescentes o pacientes, con el fin de proteger su intimidad y dignidad, como a quienes desarrollan labores de riesgo, tales como: agentes, fuentes o personal vinculado a actividades de seguridad y defensa nacional, inteligencia o contrainteligencia, para salvaguardar su vida e integridad. De igual forma, se aplica

a cualquier ciudadano, en su condición de titular de datos personales, como garantía efectiva del derecho fundamental al *Habeas Data*.

En consecuencia, deben ser objeto de anonimización todas las personas naturales identificables que aparezcan en documentos, archivos, bases de datos o registros institucionales, cuando la divulgación de su identidad no sea estrictamente necesaria para el cumplimiento de una finalidad constitucional, legal o de control público, y de manera reforzada cuando se trate de poblaciones vulnerables, víctimas, menores de edad o información asociada a contextos de riesgo.

Es importante advertir que una misma persona puede pertenecer simultáneamente a varias de las categorías descritas a continuación: una víctima puede ser menor de edad, pertenecer a una comunidad étnica y estar vinculada a un proceso judicial al mismo tiempo. En estos casos de concurrencia de categorías, el nivel de protección aplicable será el más alto de todos los que correspondan, en aplicación del principio de protección reforzada acumulativa. Este principio debe orientar el análisis contextual de riesgo que precede a toda decisión de anonimización. A continuación, se describen los principales sujetos que deben considerarse en el proceso de anonimización:

Personas naturales identificables. Cualquier persona natural cuya identidad pueda ser determinada directa o indirectamente, incluso por inferencia contextual a partir de la combinación de datos disponibles en el documento o en fuentes externas accesibles. Esta categoría es residual y comprende a toda persona que no encaje en las categorías específicas descritas a continuación.

Servidores públicos y exservidores públicos. El carácter público del cargo no elimina automáticamente la protección de datos personales del servidor. La anonimización procede cuando la identificación no es necesaria para el ejercicio del control ciudadano, cuando se trata de datos sensibles, semiprivados o de contexto personal, o cuando la identificación del servidor pueda exponer indirectamente a terceros protegidos. Son ejemplos de esta categoría los funcionarios activos y retirados, el personal administrativo, operativo y de apoyo, y los cargos directivos cuando estén involucrados menores de edad u otros sujetos de especial protección.

Víctimas. La anonimización de las víctimas constituye un deber reforzado del Estado, derivado de las obligaciones de no repetición y de garantía de no revictimización. Esta categoría comprende a las víctimas del conflicto armado, las víctimas de violaciones a derechos humanos, las víctimas de delitos comunes o graves, las víctimas de desaparición forzada y sus familiares, y las víctimas indirectas como familiares, vecinos o personas afectadas de manera colateral por los hechos documentados.

Niños, niñas y adolescentes. Esta categoría requiere protección total e irrestricta, independientemente del contexto en que aparezcan en el documento. Comprende a los menores de edad en cualquier actuación, a los menores intersexuales, y a sus representantes legales, defensores y agentes oficiosos, cuando su identificación permita inferir la identidad del menor. También deben anonimizarse los jueces, médicos tratantes u otros funcionarios cuando su identificación exponga indirectamente al menor involucrado.

Personas en condición de vulnerabilidad o especial protección. La protección se activa por la condición personal o contextual del titular, con independencia de su relación con el objeto principal del documento. Esta categoría incluye a las personas intersexuales o con ambigüedad genital, la población LGBTIQ+, las personas con VIH/SIDA o enfermedades catastróficas, las personas con discapacidad, especialmente en contextos de violencia, las mujeres adultas mayores, cuando se afecte su intimidad, las mujeres víctimas de violencia intrafamiliar, sexual o basada en género, y los migrantes, refugiados, desplazados, comunidades indígenas y familias pertenecientes a grupos étnicos.

Personas vinculadas a procesos judiciales o administrativos. La protección aplica mientras no exista sentencia en firme y, aun después, cuando subsista riesgo de afectación a derechos o cuando la información esté excluida del régimen de publicidad judicial. Comprende a los investigados, indiciados, procesados o disciplinados; a las personas mencionadas incidentalmente en expedientes; a los defensores, agentes oficiosos y representantes, cuando su identificación genere riesgo; a los ciudadanos mencionados en investigaciones penales, y a los testigos, denunciantes e informantes.

Sujetos en contextos de inteligencia, contrainteligencia y seguridad y defensa del Estado. Esta categoría comprende a los agentes de inteligencia y contrainteligencia, las fuentes humanas, los analistas, enlaces y personal de apoyo, el núcleo familiar de los agentes, y las personas mencionadas en informes de riesgo, seguridad o archivos depurados de organismos de inteligencia del Estado. La protección de estos sujetos no está condicionada a la reserva formal del documento: aun en procesos de desclasificación, su identidad debe mantenerse protegida cuando su divulgación pueda generar riesgos para su vida o integridad.

Sujetos del conflicto armado y la justicia transicional. En el marco de los mecanismos del Sistema Integral de Verdad, Justicia, Reparación y No Repetición, esta categoría requiere una protección diferenciada por contexto. Comprende a los testigos y declarantes, para quienes se recomienda el uso de códigos o seudónimos; a las víctimas de violaciones a derechos humanos;

a los terceros civiles y familiares mencionados en relatos; y a los entrevistadores y funcionarios de derechos humanos, cuando su identificación pueda generar riesgo.

Terceros no relevantes para la finalidad pública. Son personas que aparecen en el documento de manera incidental, sin relación directa con el objeto de la actuación administrativa o judicial. Su identificación no aporta valor al cumplimiento de la finalidad pública del documento y su exposición carece de justificación. Son ejemplos de esta categoría los familiares, vecinos, conocidos, testigos ocasionales y ciudadanos mencionados de manera colateral.

Pacientes y personal de salud en contextos sensibles. Esta categoría comprende a los pacientes y usuarios del sistema de salud, cuya información de diagnósticos, tratamientos e historias clínicas goza de reserva legal, conforme a la Ley 23 de 1981. También incluye a los médicos tratantes y demás personal sanitario, cuando su identificación pueda revelar indirectamente la identidad del paciente o exponer información protegida.

Personas fallecidas. Si bien podría argumentarse que el derecho a la protección de datos se extingue con la muerte, la Corte Constitucional colombiana ha reconocido que la dignidad humana y el derecho a la honra pueden protegerse *post mortem*, y que los familiares tienen legitimidad para invocar esa protección. En consecuencia, la anonimización de personas fallecidas procede cuando subsistan derechos de sus familiares que puedan verse afectados, cuando exista riesgo de estigmatización, retaliación o revictimización de personas cercanas, o cuando se trate de casos de memoria histórica en los que la exposición innecesaria de la identidad no contribuya a la finalidad de verdad y reparación.

Personas jurídicas en casos excepcionales. La regla general es que las personas jurídicas no son titulares del derecho a la protección de datos personales. Sin embargo, la anonimización puede proceder excepcionalmente cuando detrás de la identificación de la persona jurídica sea posible identificar a una persona natural, cuyos derechos fundamentales pueden verse afectados. Son ejemplos de esta categoría las razones sociales en procesos que afecten la intimidad de los accionistas o socios, y la información crediticia, financiera, regulada por la Ley 1266 de 2008.

8.2 Espacios, lugares y hechos objeto de protección

Con el fin de complementar la identificación de sujetos de protección, esta sección establece de manera sistemática los lugares, espacios físicos y hechos reales que, por su naturaleza o nivel de detalle, pueden requerir procesos de anonimización, generalización o supresión, con el fin de evitar la exposición indebida de datos personales. La correcta aplicación de estos criterios permite cumplir un principio fundamental: hacer pública la información que debe serlo, sin hacer públicas a las personas que deben ser protegidas.

Al igual que ocurre con los sujetos de protección, estos criterios deben armonizarse con el caso concreto que se esté evaluando. No constituyen reglas de aplicación automática: su procedencia depende del análisis contextual del riesgo de identificación directa o indirecta que genere cada situación específica.

Lugares e instituciones asociadas a procesos judiciales y administrativos sensibles: En esta categoría se incluyen los sitios específicos donde se vulneraron derechos, cuya mención puede permitir identificar a la persona afectada; las instituciones educativas como colegios, escuelas, jardines y universidades, cuando el asunto involucra menores de edad; las comisarías de familia en procesos relacionados con menores de edad o violencia intrafamiliar; los juzgados de familia cuando el proceso involucra niños, niñas y adolescentes; las estaciones de Policía, en casos que involucren menores de edad o víctimas de violencia; los juzgados penales, cuando la identificación del despacho permita inferir la identidad de menores o víctimas; las autoridades judiciales de instancia, cuando el asunto involucre menores de edad o mujeres víctimas de violencia; las oficinas o dependencias de atención a la mujer, incluidas las oficinas departamentales o locales; las Entidades Promotoras de Salud, cuando el asunto esté relacionado con cambios de sexo, intersexualidad u otros datos sensibles de salud; y el lugar de trabajo del titular, cuando su identificación afecte su derecho a la intimidad, al buen nombre o a la seguridad.

Identificadores administrativos y judiciales: Los números de expediente, radicados, códigos o registros de anotación, incluidos los de la Procuraduría General de la Nación u otras entidades de control, deben ser objeto de análisis, cuando su divulgación permita establecer trazabilidad personal sobre el titular, es decir, cuando a partir de ellos sea posible reconstruir la historia administrativa o judicial de una persona identificable.

Espacios físicos institucionales: En los espacios de atención y gestión administrativa deben protegerse los formularios visibles, los relatos verbales, los listados, las pantallas o tableros con datos personales en oficinas de atención al ciudadano, así como los expedientes abiertos, las

carpetas rotuladas y las actas impresas con referencias personales en despachos administrativos. En los archivos y salas de consulta pública, la protección aplica a las cajas, tomos o documentos con identificación directa o datos aún protegidos, y a los expedientes completos entregados sin supresión previa, ya sea en copia física o digital.

Espacios asociados a seguridad y control: En las instalaciones o espacios de monitoreo de seguridad, deben protegerse los reportes de incidentes, los libros de ingreso y las bitácoras con descripciones detalladas de personas. En los puestos de control y vigilancia, los listados de visitantes, los registros manuales con datos completos y las observaciones de comportamiento son elementos que pueden permitir la identificación de personas y deben ser tratados con criterio de anonimización.

Espacios vinculados a hechos reales y situaciones conflictivas: Los lugares donde ocurrieron hechos de violencia, como operativos, allanamientos o capturas, generan actas de inspección y registros fotográficos, que pueden contener datos identificadores. Las zonas afectadas por el conflicto armado, como veredas, corregimientos, barrios, mapas físicos y listados de víctimas requieren un tratamiento cuidadoso, que preserve el contexto geográfico sin singularizar a las personas. Los escenarios de protesta o movilización social también pueden generar actas, informes institucionales y registros audiovisuales que requieren anonimización antes de su divulgación.

Espacios de salud, asistencia y protección: En los centros de atención y orientación psicosocial, los registros de atención, las fichas con relatos sensibles y las agendas visibles deben ser objeto de protección. En los albergues, refugios y espacios temporales, los listados de personas atendidas, los registros de ingreso y salida y los informes de seguimiento contienen información que puede revelar la condición de vulnerabilidad de los titulares y deben ser tratados con el más alto nivel de protección.

Espacios de memoria histórica¹⁵: En museos, casas de memoria y centros de documentación, las exposiciones, relatos, fotografías, cartas o diarios que permitan identificar personas, deben ser objeto de anonimización, salvo que exista consentimiento expreso del titular o de sus familiares, o que se trate de personas públicas cuya identificación sea necesaria para la comprensión histórica del hecho. En los sitios de conmemoración y actos públicos, las placas, murales, listados

15 CNMH. (2015). Política Pública de Archivos de Derechos Humanos, Memoria Histórica y Conflicto Armado. Bogotá: Centro Nacional de Memoria Histórica. Disponible en: <https://centrodememoriahistorica.gov.co/descargas/politica-publica-archivo-ddhh.pdf>

visibles, lecturas públicas de testimonios y material impreso distribuido deben someterse al mismo criterio.

Espacios de deliberación y decisión: En las salas de reuniones y comités, las actas impresas, las grabaciones internas y las presentaciones con datos reales de personas deben ser anonimizadas antes de su archivo o divulgación. En las audiencias públicas presenciales, las intervenciones con datos de terceros, los registros audiovisuales y las transcripciones literales pueden contener información personal que no es necesaria para la comprensión pública del asunto y debe suprimirse o disociarse.

Hechos reales y situaciones fácticas susceptibles de anonimización: Los testimonios de víctimas, testigos, denunciantes y versiones libres o entrevistas constituyen declaraciones presenciales de alto riesgo de identificación y deben ser tratados con especial cuidado. Los informes de inspecciones, visitas y verificaciones pueden contener nombres de residentes o funcionarios, fotografías de espacios privados y coordenadas exactas que permiten localizar a personas. Los conflictos laborales, las denuncias internas y las investigaciones preliminares son incidentes internos que pueden afectar la honra y el buen nombre de los involucrados antes de que exista una decisión en firme.

Las violaciones a derechos humanos e infracciones al Derecho Internacional Humanitario¹⁶, como el desplazamiento forzado, la privación injusta de la libertad, la tortura, el secuestro, las ejecuciones extrajudiciales y las masacres, constituyen la categoría de mayor sensibilidad y requieren el nivel más alto de protección. Los eventos operativos de inteligencia y seguridad, incluyendo seguimientos, interceptaciones, infiltraciones y descripciones del *modus operandi*, deben anonimizarse para proteger tanto a las fuentes como a las personas objeto de las actuaciones. Los accidentes con vehículos oficiales, las fallas médicas y las estafas masivas con identificación de víctimas son accidentes y fallas del servicio, que requieren anonimización de las personas afectadas antes de su divulgación institucional.

En cuanto a los eventos históricos de alto impacto, aplica una regla especial: cuando el hecho, la fecha y el lugar son de conocimiento público (como ocurre, por ejemplo, con la Toma del

16 Archivo General de la Nación (AGN) y Centro Nacional de Memoria Histórica (CNMH). (2017). Protocolo de gestión documental de los archivos referidos a las graves y manifiestas violaciones a los Derechos Humanos e infracciones al Derecho Internacional Humanitario, ocurridas con ocasión del conflicto armado interno. Bogotá. Disponible en:

https://www.archivogeneral.gov.co/sites/default/files/Estructura_Web/5_Consulte/Recursos/Publicaciones/2022/protocolo-gd-ddhh.pdf

Palacio de Justicia), dicha información puede conservarse, pero deben anonimizarse a las personas que no tienen la condición de figuras públicas y cuya identificación no es necesaria para la comprensión histórica del evento.

Elementos contextuales que también deben anonimizarse

Además de los espacios y hechos descritos, existen elementos contextuales que, por su nivel de detalle, pueden permitir la identificación de personas y deben ser objeto de tratamiento. Entre ellos se encuentran las direcciones urbanas exactas como calles, apartamentos, edificios y conjuntos residenciales; los identificadores geográficos rurales como veredas, fincas, ríos, cerros y montañas; las coordenadas GPS exactas, que deben generalizarse o perturbarse; las subregiones altamente identificables cuando la persona es conocida localmente; los gentilicios específicos de municipios o comunidades pequeñas; y las lenguas indígenas específicas cuando la comunidad es reducida, caso en el cual debe reemplazarse la referencia por la expresión genérica "lengua indígena".

Se recomienda aplicar un criterio de granularidad como principio orientador: si el nivel de detalle del lugar, espacio o hecho permite singularizar a una persona, dicho elemento debe ser suprimido, generalizado o disociado, conservando únicamente el contexto macro, necesario para la comprensión pública del asunto. A mayor especificidad del dato geográfico, contextual o fáctico, mayor es el riesgo de identificación y más estricta debe ser la intervención de anonimización.

9. Técnicas de anonimización para información no estructurada

La anonimización en información no estructurada no se limita a una única técnica. Dependiendo del soporte, del nivel de riesgo y de la finalidad de divulgación, la entidad podrá aplicar distintas modalidades de intervención. La selección de la técnica adecuada no es una decisión discrecional: debe estar precedida de un análisis de riesgo contextual que considere la naturaleza del dato, el tipo de soporte, el universo de personas que podrían acceder al documento y la probabilidad razonable de reidentificación. A continuación, se describen las principales técnicas aplicables en documentos narrativos, audiovisuales o híbridos¹⁷.

17 Lison, P., Pilán, I., Sánchez, D., Batet, M., y Øvrelid, L. (2022). The Text Anonymization Benchmark (TAB): A Dedicated Corpus and Evaluation Framework for Text Anonymization. Computational Linguistics, MIT Press. Disponible en: <https://direct.mit.edu/coli/article/48/4/1053/112770>

9.1 Supresión o eliminación directa

La supresión o eliminación directa consiste en retirar completamente del contenido el dato identificador, sin sustituirlo por otra referencia¹⁸. Esta técnica resulta adecuada cuando la identidad del titular no es necesaria para comprender el documento, ni para cumplir una finalidad constitucional, legal o de control público. Su principal ventaja radica en su simplicidad y en la reducción inmediata del riesgo de identificación, tanto en soportes físicos como digitales. Sin embargo, puede afectar la coherencia narrativa o generar vacíos interpretativos si se eliminan múltiples referencias sin un criterio estructurado, y no neutraliza por sí sola los riesgos derivados de la singularización contextual.

Una advertencia técnica crítica en este procedimiento es la gestión de metadatos: es frecuente que el dato sea suprimido del cuerpo visible del documento, pero permanezca en las propiedades del archivo digital, como el nombre del autor, el historial de revisiones o los comentarios ocultos en archivos Word o PDF. Antes de divulgar un documento con supresión aplicada, deben eliminarse también los metadatos que puedan contener información personal. En soportes físicos, la supresión debe realizarse mediante tachado con tinta opaca e irreversible: el uso de corrector líquido, lápiz o coberturas removibles no garantiza la eliminación definitiva del dato y puede revertirse fácilmente.

Ejemplo de aplicación: En la publicación de resoluciones y actos administrativos que contienen datos semiprivados de personas naturales, como direcciones de residencia, correos electrónicos o números de teléfono personales, puede suprimirse esa información antes de la divulgación, conservando íntegro el contenido decisorio del acto. Esta práctica es coherente con el criterio operativo del artículo 31 del Decreto 103 de 2015, compilado en el Decreto 1081 de 2015 (artículo 2.1.1.1.1.) y en el Decreto 1080 de 2015 (artículo 2.8.4.3.2.), que autoriza expresamente el tachado o la edición de información en documentos con contenido mixto. En soporte digital, la supresión debe acompañarse de la eliminación de metadatos del archivo que puedan contener información identificadora, como el historial de revisiones o los comentarios ocultos; en soporte físico, debe emplearse tinta opaca permanente aplicada sobre el dato, verificando que no sea legible bajo luz ultravioleta o mediante escaneo de alta resolución.

18 Article 29 Working Party. (2014). Opinion 05/2014 on Anonymisation Techniques (WP216). European Commission. Disponible en: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf

9.2 Sustitución por etiquetas o codificación

La sustitución por etiquetas consiste en reemplazar el nombre o identificador por un código uniforme, como [FUNCIONARIO A], [TESTIGO 1] o [ACCIONANTE], que se mantiene consistente en todo el documento y sus anexos. Esta técnica es recomendable cuando resulta necesario conservar la coherencia del relato o diferenciar sujetos dentro de un expediente complejo. Su principal ventaja es que preserva la comprensión integral del documento y facilita la armonización entre distintos soportes, especialmente en expedientes judiciales, disciplinarios o administrativos extensos. Sin embargo, si el universo poblacional es reducido o si la codificación no se aplica de manera consistente en todos los documentos relacionados, puede persistir un riesgo significativo de reidentificación.

Es indispensable que la entidad elabore y conserve bajo reserva un glosario interno de códigos que establezca la correspondencia entre cada etiqueta y la identidad real del titular. Este glosario debe custodiarse con las mismas medidas de seguridad que se aplican a la información reservada y solo debe ser accesible para el personal autorizado que requiera esa correspondencia para el cumplimiento de sus funciones. Asimismo, los códigos deben ser semánticamente neutros: el uso de iniciales como [M.G.R.] puede ser suficiente para identificar a una persona en comunidades pequeñas o contextos institucionales cerrados, por lo que se recomienda el uso de códigos alfanuméricos o etiquetas genéricas que no guarden relación morfológica con el nombre real.

Ejemplo de aplicación: Es común en providencias judiciales que involucren múltiples intervinientes, donde los nombres se sustituyen por iniciales o códigos que se mantienen uniformes en todo el texto y en las versiones públicas disponibles en repositorios institucionales. Para garantizar la coherencia entre documentos de un mismo expediente, el glosario de códigos debe elaborarse desde el inicio del proceso y aplicarse de manera uniforme en todas las piezas documentales, incluidos los anexos, las pruebas y las comunicaciones asociadas. Un código que varía entre documentos del mismo expediente puede permitir la reidentificación por correlación.

9.3 Generalización

La generalización consiste en reducir el nivel de detalle de un dato para evitar la singularización, reemplazando información específica por categorías más amplias: la edad exacta por un rango etario, la dirección concreta por el municipio, o la fecha precisa por el año. Esta técnica se emplea

cuando la combinación de atributos puede permitir la identificación indirecta del titular¹⁹. Su ventaja principal es que mantiene el valor informativo global del documento, especialmente en informes narrativos o de carácter estadístico. Su aplicación exige un análisis contextual de cuidado, debido a que una generalización insuficiente puede no eliminar el riesgo de identificación, mientras que una generalización excesiva puede distorsionar el contenido y afectar su valor probatorio o histórico, y afectar el derecho de acceso y transparencia a la información pública.

Como orientación práctica, puede aplicarse el siguiente esquema de niveles de generalización progresiva por tipo de dato:

- Edad: edad exacta → rango de diez años → grupo etario (menor, adulto joven, adulto, adulto mayor)
- Ubicación geográfica urbana: dirección exacta → barrio → localidad → municipio → departamento
- Ubicación geográfica rural: nombre de finca o vereda → corregimiento → municipio → subregión
- Fecha: fecha exacta → mes y año → año → período (primer semestre, segundo semestre).

La selección del nivel de generalización adecuado depende del tamaño de la población de referencia: en municipios o comunidades pequeñas, incluso datos aparentemente generales como "mujer adulta mayor indígena desplazada" pueden ser suficientes para identificar a alguien, si el universo de personas con esas características es reducido. En estos casos, debe aplicarse un nivel de generalización adicional o combinarse con supresión.

Ejemplo de aplicación: En informes sobre violencia basada en género, puede reemplazarse la expresión "mujer de 47 años residente en el barrio X" por "mujer adulta residente en el municipio", cuando el barrio específico permita la identificación en poblaciones pequeñas o comunidades conocidas localmente. Si el municipio también es pequeño y la combinación de atributos sigue siendo identificadora, debe generalizarse al nivel departamental o suprimirse el atributo de menor relevancia para la finalidad del informe.

19 Rocher, L., Hendrickx, J., & de Montjoye, Y. A. (2019). Estimating the success of re-identifications in incomplete datasets using generative models. Nature Communications. Disponible en: <https://www.nature.com/articles/s41467-019-10933-3>

9.4 Disociación irreversible

La disociación irreversible tiene como finalidad eliminar cualquier posibilidad razonable de vincular el dato con una persona determinada, incluso mediante cruces con fuentes externas. Se utiliza en contextos de alto riesgo, publicación masiva o cuando la información puede permanecer disponible por largos periodos, como ocurre en archivos históricos digitalizados. Su principal fortaleza es que ofrece el máximo nivel de protección frente a la reidentificación futura. No obstante, es la técnica que mayor impacto tiene sobre el valor narrativo, probatorio e histórico del documento, razón por la cual su aplicación debe estar precedida de un análisis profundo y documentado que justifique su procedencia frente a técnicas menos invasivas.

Antes de considerar un conjunto de datos como efectivamente disociado, la entidad debe aplicar un test de reidentificación, que consiste en verificar si es posible vincular al titular cruzando el documento con fuentes abiertas accesibles, como redes sociales, noticias de prensa, registros públicos, bases de datos institucionales o motores de búsqueda. Si el cruce permite la identificación, la disociación no es suficiente y debe reforzarse con supresión o generalización adicional.

Una especial consideración se debe tener en archivos históricos digitalizados con reconocimiento óptico de caracteres (OCR): el texto resultante es indexable por motores de búsqueda, lo que amplía significativamente el riesgo de reidentificación. Un nombre que en un documento físico de archivo solo podía ser encontrado por quien tuviera acceso físico al expediente, en formato digital con OCR puede ser localizado en segundos mediante una búsqueda en internet. Este factor debe considerarse de manera específica en los procesos de digitalización masiva de acervos documentales.

Ejemplo de aplicación: Es pertinente en procesos de digitalización de archivos relacionados con derechos humanos o conflicto armado, cuando la exposición pública podría generar revictimización o riesgos para familiares de las personas mencionadas. En estos casos, el test de reidentificación debe realizarse no solo sobre el documento aislado sino sobre el conjunto del acervo digitalizado, verificando que la combinación de documentos disponibles no permita reconstruir la identidad del titular a partir de referencias dispersas.

9.5 Distorsión técnica en formatos audiovisuales

La distorsión técnica implica la modificación del contenido visual o sonoro para impedir el reconocimiento de la persona, mediante pixelado robusto, cobertura opaca, desenfoque irreversible o alteración biométrica de la voz²⁰. Esta técnica es aplicable en audios, videos, grabaciones de audiencias y material gráfico institucional. Es indispensable que la intervención técnica sea robusta e irreversible: distorsiones superficiales o de baja resolución pueden revertirse mediante herramientas digitales de mejora de imagen o voz, lo que las convierte en ineficaces y puede generar una falsa sensación de protección.

En cuanto al tratamiento de imágenes y video, existe una diferencia técnica relevante entre el pixelado estático y el pixelado dinámico con seguimiento de movimiento. El pixelado estático aplica la distorsión sobre una región fija del fotograma y puede dejar expuesto el rostro cuando la persona se desplaza dentro del encuadre. El pixelado dinámico con seguimiento de movimiento rastrea el rostro u objeto a lo largo de toda la grabación y es técnicamente más robusto. Adicionalmente, herramientas de inteligencia artificial de mejora de imagen pueden revertir pixelados de baja intensidad, por lo que la distorsión debe aplicarse con la mayor intensidad técnicamente posible.

Con respecto al tratamiento de audio, no basta con distorsionar la voz del hablante principal si en el registro sonoro pueden identificarse otros elementos como nombres mencionados en el fondo, referencias a lugares específicos, sonidos institucionales reconocibles como el himno de un colegio o las campanadas de una iglesia o acentos y dialectos que singularicen al hablante. Todos estos elementos deben ser identificados y tratados antes de la divulgación del registro.

Finalmente, los archivos de video y audio contienen metadatos que pueden incluir la ubicación GPS del lugar de grabación, la fecha y hora exactas, el nombre del dispositivo utilizado y el nombre del usuario registrado en el equipo. Estos metadatos deben eliminarse antes de la divulgación, independientemente de las distorsiones aplicadas al contenido visual o sonoro.

20 Kleinhans, B., y otros. (2024). Fantômas: Understanding Face Anonymization Reversibility. Disponible en: <https://www.researchgate.net/publication/239665775>

Ejemplo de aplicación: En la difusión de videos institucionales sobre operativos judiciales o audiencias públicas, puede distorsionarse la voz de testigos y aplicarse seguimiento dinámico de rostros mediante pixelado. Adicionalmente, debe verificarse que el audio de fondo no contenga nombres ni referencias identificadoras, y que los metadatos del archivo hayan sido eliminados antes de su publicación o entrega.

9.6 Seudonimización controlada

Laseudonimización consiste en sustituir la identidad real por un código, conservando internamente una tabla de correspondencia bajo estrictas medidas de custodia y acceso restringido. Esta técnica no constituye anonimización en sentido estricto, ya que el proceso puede revertirse si se accede a la tabla de equivalencias²¹. Por esta razón, no debe utilizarse como técnica de anonimización definitiva para divulgación pública, sino como mecanismo de gestión interna en procesos donde se requiere trazabilidad institucional controlada, como investigaciones, análisis histórico o proyectos de interoperabilidad. Su aplicación exige protocolos estrictos de seguridad, custodia y auditoría de acceso a la tabla de correspondencias.

La tabla de equivalencias entre seudónimos e identidades reales debe cumplir como mínimo los siguientes requisitos de seguridad: acceso restringido a un número mínimo de personas expresamente autorizadas, preferiblemente no más de dos; cifrado del archivo que la contiene; registro de auditoría de cada acceso realizado; y protocolo de destrucción segura, cuando la tabla ya no sea necesaria para los fines que motivaron su creación. La ausencia de cualquiera de estos controles convierte laseudonimización en una medida insuficiente.

Adicionalmente, debe tenerse en cuenta que laseudonimización no protege contra ataques de reidentificación por correlación: si un mismo seudónimo aparece en múltiples documentos con atributos consistentes como edad, cargo, municipio de residencia o condición de salud, la combinación de esos atributos puede ser suficiente para identificar al titular, sin necesidad de acceder a la tabla de equivalencias. Por esta razón, laseudonimización debe combinarse con generalización de los atributos contextuales, cuando el riesgo de correlación sea significativo.

21 Autoridade Nacional de Proteção de Dados (ANPD) de Brasil. (2024). Estudio Preliminar sobre Anonimización y Seudonimización. Disponible en: <https://www.gov.br/anpd>

Ejemplo de aplicación: En proyectos académicos o de memoria histórica coordinados por entidades públicas, los nombres reales pueden mantenerse bajo reserva institucional mientras la versión pública se divulga con códigos. La tabla de equivalencias debe custodiarse con cifrado, acceso restringido y auditoría de accesos, y debe establecerse desde el inicio del proyecto un protocolo claro sobre quién puede acceder a ella, en qué circunstancias y qué ocurre con ella al finalizar el proyecto.

9.7 Acceso parcial o mediado

El acceso parcial o mediado no implica modificar el documento original, sino limitar la información que se entrega, proporcionando únicamente los datos estrictamente necesarios para satisfacer la solicitud o la finalidad pública. Esta técnica es adecuada cuando la anonimización total afecta la comprensión sustantiva del documento o cuando existe riesgo elevado de reidentificación mediante el acceso al expediente completo. Su aplicación debe estar motivada de manera expresa y documentada, indicando qué información se entrega, qué información se retiene y con qué fundamento normativo.

Desde el punto de vista operativo, el acceso parcial puede implementarse de dos maneras. La primera es el acceso mediado presencial, que consiste en habilitar la consulta del documento en sala de archivo bajo condiciones controladas, con restricciones sobre la reproducción, fotografía o copia de las partes protegidas, y con acompañamiento de un servidor público que oriente la consulta dentro de los límites autorizados. La segunda es el acceso mediado digital, que consiste en entregar extractos, respuestas puntuales o versiones parciales del documento sin remitir el expediente completo, cuando este contiene información sensible de terceros que no guarda relación directa con la solicitud.

En ambos casos, la entidad debe documentar en el expediente qué información fue entregada o consultada, a quién, en qué fecha y con qué fundamento normativo. Esta documentación constituye una garantía de trazabilidad institucional y permite demostrar, ante eventuales reclamaciones, que el acceso se otorgó dentro de los límites legalmente establecidos.

Ejemplo de aplicación: Ante una solicitud de acceso a un expediente administrativo, puede responderse con el dato puntual requerido por el solicitante sin entregar la totalidad del expediente, cuando este contiene información sensible de terceros que no guarda relación directa con la solicitud. La respuesta debe indicar expresamente qué información se entrega, qué

información se retiene y el fundamento normativo de esa retención, dejando constancia de ello tanto en la comunicación al solicitante, como en el expediente interno de la entidad.

9.8 Combinación de técnicas

Las técnicas de anonimización en información no estructurada rara vez se aplican de manera aislada. Con frecuencia es necesario combinar supresión y generalización, codificación y distorsión técnica, o acceso parcial con disociación contextual. La combinación de técnicas no es acumulativa en sentido automático: cada técnica debe seleccionarse en función del tipo de dato, del soporte y del riesgo específico que se busca neutralizar. Una combinación inadecuada puede generar tanto sobreprotección que afecte el valor del documento, como subprotección, que deje subsistentes riesgos de reidentificación. La selección de la técnica o combinación de técnicas deberá basarse siempre en el análisis de riesgo contextual previamente realizado, en la naturaleza del soporte y en el alcance de la divulgación prevista, asegurando que la probabilidad razonable de reidentificación sea baja y proporcional al interés público comprometido.

10. Riesgos de reidentificación

El objetivo central de la anonimización es impedir de manera irreversible que una persona pueda ser identificada a partir de los datos tratados, ya sea de forma directa o mediante el cruce con fuentes externas. Cuando este objetivo no se alcanza plenamente, no se está ante una anonimización propiamente dicha, sino ante una seudonimización, en la que la identificación sigue siendo posible para quien disponga de los elementos necesarios para realizarla.

Esta distinción tiene implicaciones jurídicas y operativas relevantes. Desde la perspectiva de las entidades públicas, que gestionan la información y conocen su contexto de producción, la protección aplicada puede ser técnicamente una seudonimización, puesto que la entidad conserva los elementos que permitirían reidentificar al titular. Sin embargo, para terceros que no tienen acceso a esos elementos, la información puede considerarse anonimizada, en la medida en que la reidentificación sea técnicamente difícil o prácticamente imposible con los medios razonablemente disponibles. Esta diferencia entre el umbral interno y el umbral externo de protección debe ser considerada explícitamente en el diseño de los procesos de anonimización.

Los riesgos de reidentificación constituyen uno de los principales desafíos técnicos de la anonimización, y se incrementan de manera constante como consecuencia del avance tecnológico, la proliferación de fuentes abiertas de información y el desarrollo de herramientas automatizadas

de extracción y correlación de datos. Entre estas herramientas se encuentran el raspado de datos (*scraping*), los rastreadores automatizados (*crawlers*), los robots de indexación y los grandes modelos de lenguaje, que pueden procesar, asociar y recuperar información personal a una velocidad y escala que superan ampliamente las capacidades de revisión manual. Este contexto tecnológico obliga a las entidades públicas a adoptar una perspectiva dinámica frente al riesgo de reidentificación: una técnica de anonimización que hoy es suficiente puede dejar de serlo en el futuro, razón por la cual los procesos de anonimización deben revisarse periódicamente a la luz del estado de la técnica disponible.

10.1 Modalidades técnicas de riesgo de reidentificación

Los riesgos de reidentificación no se presentan de manera uniforme. La información técnica y la práctica internacional identifican tres modalidades principales²², cuya comprensión es indispensable para diseñar procesos de anonimización efectivos:

TÉCNICA	DESCRIPCIÓN
Singularización	Es la posibilidad de que una persona pueda ser identificada a partir de datos anonimizados, ya sea de manera directa o mediante técnicas de desanonimización, lo cual se puede presentar cuando los datos anonimizados no están lo suficientemente protegidos o cuando se pueden vincular con fuentes de información externas ²³ . En esta técnica el riesgo es más directo e inmediato, basta con que una combinación de atributos sea única dentro del universo de referencia para que la singularización sea posible, aunque ninguno de los atributos individualmente considerados sea un identificador directo
Vinculabilidad	Es la capacidad de vincular o conectar datos anonimizados con datos externos u otras fuentes de información, lo que podría llevar a la identificación de grupos. Este riesgo se incrementa en entornos de fuentes abiertas al público (ejemplo: redes sociales, noticias, bases de datos abiertas), en el cual hay una alta posibilidad y capacidad de cruce de fuentes de información

22 Article 29 Working Party. (2014). Opinion 05/2014 on Anonymisation Techniques (WP216). European Commission. Disponible en: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf

23 Rocher, L., Hendrickx, J., & de Montjoye, Y. A. (2019). Estimating the success of re-identifications in incomplete datasets using generative models. Nature Communications. <https://www.nature.com/articles/s41467-019-10933-3>

TÉCNICA	DESCRIPCIÓN
Inferencia	Es la capacidad de deducir información sensible o identificar a personas, a partir de datos anonimizados, utilizando lógica o razonamiento ²⁴ . Este tipo de técnica no depende necesariamente de fuentes abiertas o accesibles; los mismos datos e información del documento, incluso datos anonimizados pueden deducirse la información que no se ha suministrado de manera directa, pero que, a través de análisis de patrones de la información en su conjunto se llega a esta

Fuente: elaboración propia.

Los avances recientes en inteligencia artificial han transformado el panorama del riesgo de reidentificación, especialmente en documentos que contienen texto libre, relatos o descripciones. Los Modelos de Lenguaje de Gran Escala (LLM) son capaces de identificar a una persona a partir de un documento aparentemente anonimizado, sin necesidad de contar con sus datos directos como nombre o número de identificación (Lermen et al., 2026). Estos modelos leen el texto completo y extraen pequeños detalles que por separado parecen inofensivos, como puede ser la mención de un colega, una ciudad, un pasatiempo o una anécdota laboral, y los combinan para construir un perfil que luego contrastan con información disponible en internet, como hojas de vida, perfiles profesionales o publicaciones en redes sociales.

Frente a las tres modalidades de riesgo descritas en la tabla, el impacto es directo: en la singularización, porque unen detalles dispersos en el texto para construir una huella de identidad única sin identificadores directos; en la vinculabilidad, porque conectan el contenido del documento con fuentes públicas de manera automática y a muy bajo costo, eliminando la protección que antes ofrecía la dificultad práctica de cruzar esas fuentes manualmente; y en la inferencia, porque no solo detectan patrones sino que razonan sobre la evidencia disponible, deduciendo atributos sensibles que nunca fueron mencionados explícitamente. En pruebas controladas, este tipo de análisis logró identificar correctamente a casi 7 de cada 10 personas en documentos aparentemente anonimizados, a un costo de entre uno y cuatro dólares por persona (Lermen et al., 2026).

Esto significa que eliminar los identificadores directos de un documento ya no es suficiente para garantizar el anonimato de quienes aparecen en él, y que las técnicas tradicionales

24 Staab, R., Vero, M., Balunović, M., & Vechev, M. (2024). Beyond Memorization: Violating Privacy Via Inference with Large Language Models. ICLR 2024. <https://arxiv.org/abs/2310.07298>

diseñadas para bases de datos estructuradas no fueron pensadas para enfrentar este tipo de análisis sobre texto libre (Lermen et al., 2026). Por esta razón, antes de divulgar o transferir un documento anonimizado, es recomendable evaluar si un sistema de inteligencia artificial podría, a partir de los detalles que quedaron en el texto, reconstruir la identidad de las personas mencionadas cruzando esa información con fuentes abiertas, con el fin de identificar qué combinaciones de datos requieren una intervención adicional antes de la divulgación (Lermen et al., 2026).²⁵

10.2 Inteligencia artificial y los retos de reidentificación

El desarrollo acelerado de la inteligencia artificial introduce riesgos distintos a los tradicionalmente considerados en los procesos de anonimización. Los modelos de lenguaje de gran escala, los algoritmos de inferencia estadística, han demostrado capacidad para reidentificar personas, a partir de conjuntos de datos que fueron considerados suficientemente anonimizados. Esta capacidad no depende necesariamente del acceso a identificadores directos, debido a que los modelos de inteligencia artificial pueden inferir la identidad de una persona a partir de patrones de comportamiento, combinaciones de atributos aparentemente neutros, estilos de escritura, características vocales residuales o elementos contextuales que un análisis humano convencional no detectaría²⁶.

En el ámbito de los documentos de archivo, esto implica que textos narrativos anonimizados pueden ser procesados por modelos de lenguaje, para inferir la identidad del autor o del sujeto a partir del estilo en la redacción, el vocabulario empleado o las referencias contextuales. Las imágenes con rostros distorsionados pueden ser sometidas a técnicas de mejora basadas en inteligencia artificial que reconstruyan parcialmente los rasgos suprimidos. Los registros de voz con distorsión biométrica pueden ser procesados por modelos de separación de fuentes sonoras que recuperen características identificadoras. Y los datos aparentemente disociados pueden ser correlacionados automáticamente con fuentes abiertas, mediante técnicas de aprendizaje automático que identifiquen coincidencias no evidentes para el análisis humano²⁷.

25 Lermen, S., Paleka, D., Swanson, J., Aerni, M., Carlini, N., & Tramèr, F. (2026). Large-scale online deanonymization with LLMs. arXiv preprint arXiv:2602.16800. <https://arxiv.org/abs/2602.16800>

26 Staab, R., Vero, M., Balunović, M., & Vechev, M. (2024). Beyond Memorization: Violating Privacy Via Inference with Large Language Models. The Twelfth International Conference on Learning Representations (ICLR 2024). Disponible en: <https://arxiv.org/abs/2310.07298>

27 Wiepert, D., et al. (2024). Reidentification of Participants in Shared Clinical Data Sets: Experimental Study. JMIR AI. Disponible en: <https://ai.jmir.org/2024/1/e52054>

Adicionalmente, los grandes modelos de lenguaje se entrenan con volúmenes masivos de información extraída de fuentes abiertas de internet, lo que significa que información personal que fue publicada en algún momento, aunque posteriormente haya sido retirada o anonimizada, puede haber quedado incorporada en los parámetros de estos modelos y ser recuperable mediante técnicas de extracción de información de entrenamiento. Este fenómeno, conocido como “memorización de datos de entrenamiento” representa un riesgo emergente que las entidades públicas deben considerar al evaluar el impacto a largo plazo de sus decisiones de publicación²⁸.

La evaluación del riesgo de reidentificación no puede realizarse únicamente en el momento de la intervención: debe incluir una proyección razonable sobre las capacidades tecnológicas previsibles en el horizonte de tiempo durante el cual la información permanecerá disponible, especialmente cuando se trata de archivos históricos o información de divulgación permanente.

10.3 Consecuencias de una reidentificación

Una inadecuada anonimización puede generar impactos significativos en los sujetos involucrados, al comprometer derechos fundamentales y bienes jurídicos de alta relevancia constitucional. Entre estos efectos se encuentran la afectación del buen nombre, la autonomía personal y la autodeterminación informativa; la vulneración de la dignidad humana, la igualdad y la libertad de expresión; así como la exposición a escenarios de discriminación, difamación o estigmatización. Asimismo, puede incidir negativamente en derechos laborales y en el acceso efectivo a servicios, derivando incluso en la pérdida del empleo, reportes injustificados en bases de datos de morosos o incumplimientos de obligaciones materiales con perjuicios económicos. En contextos más sensibles, una deficiente protección de la identidad puede contribuir al deterioro del estado de salud, al padecimiento psicológico e incluso a la exposición a agresiones físicas. En términos estructurales, estos riesgos no solo impactan derechos individuales, sino también principios como la justicia social, los derechos humanos y los valores democráticos, al erosionar la confianza en las instituciones encargadas de custodiar y tratar la información.

28 Survey on Training Data Extraction From Pre-trained Language Models. (2023). Association for Computational Linguistics. Disponible en: <https://aclanthology.org/2023.trustnlp-1.23.pdf>

SEGUNDA PARTE

Lineamientos específicos para la anonimización y protección de datos personales en la gestión documental

- **11** Enfoque de anonimización según el ciclo de vida documental
- **12** Lineamientos de anonimización en archivo histórico
- **13** Lineamientos metodológicos en el proceso de anonimización
- **14** Lineamientos para la selección de técnicas de anonimización
- **15** Análisis de riesgo contextual y probabilidad de reidentificación
- **16** Estructura institucional y responsabilidades

11. Enfoque de anonimización según el ciclo de vida documental

El artículo 23 de la Ley 594 del 2000 establece que, teniendo en cuenta el ciclo vital de los documentos, los archivos se clasifican en: archivo de gestión, archivo central y archivo histórico. Este artículo, los define como:

- **Archivo de gestión:** Comprende toda la documentación que es sometida a continua utilización y consulta administrativa por las oficinas productoras u otras que la soliciten. Su circulación o trámite se realiza para dar respuesta o solución a los asuntos iniciados
- **Archivo central:** En el que se agrupan documentos transferidos por los distintos archivos de gestión de la entidad respectiva, cuya consulta no es tan frecuente, pero que siguen teniendo vigencia y son objeto de consulta por las propias oficinas y particulares en general
- **Archivo histórico:** Es aquel al que se transfieren desde el archivo central los documentos de archivo de conservación permanente.

En materia de anonimización, es fundamental analizar el ciclo de vida de los documentos, puesto que el régimen jurídico no es inmutable, sino que puede variar según la fase en la que se encuentren. El paso de una etapa a otra, por ejemplo, de archivo de gestión a archivo histórico, puede implicar una transformación en las condiciones de acceso. En consecuencia, la información puede estar sujeta a restricciones, adquirir carácter reservado o, en el caso de los archivos históricos, ser de acceso público, eventualmente previa aplicación de procesos de anonimización, cuando resulte necesario para la protección de derechos o datos personales.

Por otra parte, el nivel de riesgo asociado al documento contenido en el archivo no es el mismo en todas sus fases. Dicho riesgo varía según el contexto en el que se encuentre el documento: durante un trámite interno, una vez archivado, al ser puesto a disposición para consulta ciudadana, al momento de su publicación o cuando adquiere la condición de archivo histórico. En virtud de esta variabilidad, no resulta jurídicamente adecuado aplicar una regla uniforme de anonimización, pues las medidas deben ser proporcionales al nivel de exposición y a las condiciones específicas de acceso en cada fase.

Sin perjuicio de lo anterior, los procesos de anonimización no pueden desligarse de los lineamientos de gestión documental definidos en la normatividad colombiana. En consecuencia, los procesos de planeación documental, producción documental, gestión y trámite, organización documental, transferencias documentales, disposición, preservación a largo plazo y valoración

documental podrán diseñarse, gestionarse e implementarse incorporando criterios de anonimización, en la medida necesaria para garantizar el acceso a la información contenida en los archivos, sin comprometer derechos fundamentales.

En esa línea, aunque no resulte jurídicamente viable establecer una regla general de anonimización aplicable de manera uniforme a todas las fases del ciclo vital, sí es posible (y desde la perspectiva archivística, altamente recomendable) incorporar esta variable desde la planeación de la gestión documental institucional, con al menos los siguientes componentes:

- **Planeación técnica:** definir criterios para identificar documentos con información susceptible de anonimización, roles y responsabilidades, análisis de usuarios y escenarios de acceso futuro. Esta planeación debe articularse con el PINAR y el PGD y puede materializarse como un componente o programa de anonimización dentro del PGD o integrado a los programas ya existentes
- **Producción documental:** diseñar formatos y estructuras documentales eficientes y normalizadas, que reduzcan la exposición innecesaria de datos personales, y prever desde el diseño la generación de versiones de consulta cuando corresponda
- **Gestión y trámite:** incorporar la anonimización como requisito previo cuando el documento deba ser reproducido, consultado o publicado hacia terceros; incluir metadatos que permitan identificar qué documentos requieren anonimización y garantizar la trazabilidad entre el original y las versiones públicas
- **Organización documental:** durante la clasificación, ordenación y descripción, identificar presencia de datos sensibles o de especial protección para facilitar decisiones de acceso y, cuando proceda, la construcción de versiones públicas incluso de instrumentos de consulta, sin afectar el original
- **Transferencias documentales:** en las transferencias primarias y secundarias, identificar información sensible y definir la creación de copias administrativas o copias históricas para consulta. En transferencias secundarias se deberán prever lineamientos para evitar que al archivo histórico lleguen documentos cuya información no pueda ponerse a consulta pública bajo parámetros responsables
- **Disposición final:** precisar que la anonimización no reemplaza ni modifica TRD o TVD, ni constituye una decisión de disposición final
- **Preservación a largo plazo:** la intervención se realiza sobre copias de trabajo; sin embargo, la preservación debe contemplarse a largo plazo del máster, entendido como la copia de trabajo y/o versión íntegra bajo control archivístico, junto con sus metadatos y trazabilidad

- **Valoración documental:** incorporar el análisis de contexto y de clasificación (pública, reservada o clasificada), así como términos de reserva. No puede considerarse la posibilidad de que un documento reservado pase a la fase histórica; y cuando se trate de información clasificada, debe asumirse que puede sostenerse en el tiempo al no contar con plazo general predeterminado.

De este modo, la anonimización deja de concebirse como medida reactiva aplicada al final y pasa a entenderse como una variable planificada del sistema de gestión documental, sin afectar la autenticidad, integridad ni valor probatorio del documento original.

11.1 Lineamiento de anonimización en archivos de gestión

Dado que el archivo de gestión soporta la operación institucional y la resolución de trámites, no es una fase orientada a anonimizar documentos de forma ordinaria. La medida principal en esta etapa es la gestión de accesos (roles, perfiles, controles y medidas de seguridad físicas y tecnológicas), de acuerdo con el soporte o formato en que repose el documento, garantizando confidencialidad y circulación restringida.

En consecuencia, en el archivo de gestión la anonimización constituye una medida excepcional, aplicable únicamente cuando se genere una copia del documento destinada a terceros, ajenos a la dependencia o a la entidad. En esta fase, el documento cumple una finalidad operativa directa, por lo que anonimizar el original limitaría su utilidad administrativa, jurídica y funcional. Por tanto, el documento original debe conservarse íntegro y bajo condiciones que garanticen su confidencialidad, seguridad y acceso restringido. Para ello, se recomienda a la entidad:

- Restringir el acceso a los documentos con base en roles y perfiles autorizados, conforme a la Política de Seguridad de la Información y los controles de acceso definidos por la entidad
- Abstenerse de anonimizar, modificar o intervenir el documento original, con el fin de preservar su integridad, autenticidad y valor probatorio
- Garantizar la integridad documental, evitando cualquier alteración, supresión o modificación no autorizada del contenido
- Mantener mecanismos de registro y trazabilidad de accesos a la información, que permitan identificar consultas, usos o intervenciones realizadas sobre los documentos

- Aplicar de manera permanente los principios de confidencialidad, seguridad y circulación restringida de la información, conforme a lo establecido en el artículo 4 de la Ley 1581 de 2012
- Cuando sea necesario entregar o divulgar el contenido del documento a terceros, generar una copia de trabajo y evaluar la necesidad de aplicar técnicas de anonimización antes de su entrega, asegurando trazabilidad entre el original y la versión entregada
- Incorporar, desde esta etapa, metadatos que indiquen si el documento contiene información susceptible de anonimización, con el fin de facilitar decisiones en fases posteriores (archivo central e histórico).

11.2 Lineamiento de anonimización en archivo central

En la fase de archivo central, los documentos han perdido su uso operativo frecuente, pero conservan valor administrativo, jurídico, o probatorio, y pueden ser objeto de consulta ocasional o requerimientos de acceso por parte de terceros. En esta etapa, la entidad debe garantizar la conservación íntegra del documento original y, simultáneamente, preparar las condiciones necesarias para su divulgación controlada cuando exista una solicitud legítima de acceso.

En este contexto, la anonimización adquiere un carácter preventivo y necesario como medida para proteger los derechos fundamentales de las personas, sin afectar el acceso a la información pública cuando este sea procedente. En consecuencia, se sugiere que la entidad aplique las siguientes reglas:

- Conservar el documento original íntegro, sin modificaciones, alteraciones o intervenciones que afecten su integridad, autenticidad o valor probatorio
- Evaluar la necesidad de aplicar técnicas de anonimización cuando se genere una copia del documento con destino a terceros, incluidos ciudadanos, otras entidades o instancias externas autorizadas
- Analizar previamente el contenido del documento, con el fin de identificar la presencia de datos personales sensibles, datos personales de terceros que no son objeto directo del documento, datos de víctimas o sujetos de especial protección constitucional, identificadores directos o indirectos que puedan permitir la identificación de una persona, de manera individual o mediante la combinación de información contextual
- Generar, cuando corresponda, una versión pública anonimizada, que permita el acceso al contenido del documento sin comprometer la identidad de las personas, ni vulnerar sus derechos fundamentales

- Garantizar que el proceso de anonimización se realice sobre una copia de trabajo, preservando en todo momento el documento original bajo condiciones de acceso restringido.

Estas medidas permiten a la entidad cumplir tanto con las obligaciones en materia de conservación documental, acceso a la información pública y protección de datos personales, conforme a lo establecido en la Ley 1581 de 2012, la Ley 1712 de 2014 y la Ley 594 de 2000.

Sin perjuicio de lo anterior, la entidad deberá distinguir entre información clasificada e información reservada, conforme a los artículos 18 y 19 de la Ley 1712 de 2014, considerando lo siguiente:

- Los documentos sujetos a reserva deberán permanecer en archivo central sin divulgación y sin generación de versiones públicas mientras la reserva esté vigente. La valoración documental debe contemplar los términos de reserva con miras a proponer tiempos de retención en gestión y central. De ninguna manera puede considerarse que documentación reservada pase al archivo histórico
- Los documentos clasificados mantendrán su restricción mientras subsistan las razones constitucionales que la justifican; al no existir un plazo general predeterminado, su verificación debe ser periódica por el responsable del archivo, articulada con procesos de valoración.

12. Lineamiento de anonimización en archivo histórico

En la fase de archivo histórico, los documentos adquieren valor permanente por su relevancia para la memoria institucional, la investigación, la cultura y la garantía del derecho de acceso a la información pública. En esta etapa, el acceso público constituye la regla general; sin embargo, dicho acceso debe ejercerse en armonía con la protección de derechos fundamentales, en particular el derecho a la intimidad, al *Habeas Data*, al buen nombre y a la dignidad humana.

En este contexto, la anonimización es un mecanismo necesario para equilibrar el principio de publicidad de la información pública con el deber de protección de datos personales. La divulgación de documentos históricos no implica que la identidad de las personas deba hacerse pública de forma irrestricta. Cuando el contenido incluya datos personales, y especialmente datos

sensibles, la entidad deberá adoptar medidas orientadas a suprimir, disociar o limitar los elementos que permitan la identificación de los titulares.

En efecto, el literal e) del artículo 6 de la Ley 1581 de 2012 establece que el tratamiento de datos sensibles está prohibido, salvo, entre otros casos, cuando tenga finalidad histórica, estadística o científica, evento en el cual deberán adoptarse medidas conducentes a la supresión de la identidad de los titulares. Esta disposición impone el deber de implementar mecanismos de anonimización, cuando los documentos históricos contengan información sensible. Para tal efecto, la entidad deberá considerar las siguientes reglas:

- La anonimización deberá realizarse sobre copias de trabajo o versiones destinadas a consulta pública, preservando íntegramente el documento original, conforme a los principios de integridad documental, conservación y valor probatorio
- La anonimización debe generar versiones de consulta y registros técnicos de intervención, incluyendo control de cambios y trazabilidad, de modo que se asegure la relación verificable entre el original y las versiones públicas
- Generar versiones públicas anonimizadas, cuando los documentos contengan datos personales cuya divulgación pueda afectar derechos fundamentales
- Eliminar o suprimir identificadores directos (nombres, identificación, direcciones, firmas, imágenes u otros elementos), que permitan la identificación inmediata cuando ello afecte derechos fundamentales
- Evaluar y tratar identificadores indirectos o contextuales, cuya combinación permita la identificación razonable de los titulares
- Aplicar técnicas de anonimización contextual considerando el contenido y el contexto histórico, territorial, institucional y narrativo del documento
- Adoptar medidas reforzadas cuando exista información relacionada con víctimas, niños, niñas y adolescentes, salud, datos sensibles o sujetos de especial protección constitucional
- Evaluar el riesgo de reidentificación antes de autorizar el acceso o divulgación, considerando identificadores directos e indirectos.

Adicionalmente, y en consistencia con los procesos de gestión documental, el archivo histórico debe operar articulado con transferencia y preservación:

- En transferencias secundarias deben existir lineamientos para que al archivo histórico no lleguen documentos sin evaluación sobre posibilidad de consulta pública y necesidad de versiones públicas.

- La preservación a largo plazo debe contemplar la preservación del máster y de las versiones de consulta, con sus metadatos y evidencia de integridad y trazabilidad.

Finalmente, desde una perspectiva sistemática del derecho de acceso a la información pública y del principio de máxima publicidad, se debe diferenciar el tratamiento archivístico de información clasificada y reservada:

- La anonimización se proyecta principalmente respecto de información clasificada, en tanto su restricción puede sostenerse mientras subsistan razones constitucionales que la justifican y no cuenta con plazo general predeterminado
- La información reservada es temporal y se gestiona en etapas de archivo de gestión y central, mediante valoración documental y control de términos. Conforme a los criterios archivísticos, no puede considerarse que un documento reservado pase a archivo histórico.

En consecuencia, en el archivo histórico debe prevalecer el acceso a la información pública, con la única excepción de aquella que conserve, legítimamente, el carácter de clasificada; cuando corresponda, el acceso se garantizará mediante versiones públicas anonimizadas, preservando el original y la trazabilidad.

13. Lineamientos metodológicos en el proceso de anonimización

La anonimización de información contenida en documentos de archivo no constituye una actuación aislada ni una decisión discrecional adoptada caso a caso sin criterios definidos. Por el contrario, se trata de un proceso técnico-jurídico estructurado, que debe aplicarse de manera sistemática, documentada y coherente con el régimen de protección de datos personales, el derecho de acceso a la información pública y los principios de la función archivística.

En las entidades públicas la anonimización no puede reducirse a la supresión mecánica de nombres o identificadores visibles. Su adecuada aplicación exige una secuencia metodológica, que permita identificar la información susceptible de intervención, determinar el régimen jurídico aplicable, ejecutar la técnica correspondiente según el soporte documental, evaluar el riesgo de reidentificación y, finalmente, generar una versión pública que garantice simultáneamente la transparencia y la protección de los derechos fundamentales.

Los presentes lineamientos establecen los pasos, fases o aspectos que las entidades públicas deben integrar en el proceso institucional de anonimización. Estas fases deberán aplicarse de manera proporcional, atendiendo a la naturaleza del documento, el contexto de divulgación y el nivel de riesgo asociado, asegurando en todo momento la preservación del documento original y la integridad del acervo documental.

13.1 Identificación de información susceptible de anonimización

La identificación de información susceptible de anonimización constituye la fase inicial de la anonimización. Esta fase no se ejecuta de manera abstracta ni generalizada, sino en función del escenario que activa el proceso. La entidad deberá determinar previamente cuál es el supuesto que origina la necesidad de anonimizar, pueden ser, entre otros:

- Solicitud específica de acceso a la información (transparencia pasiva)
- Publicación obligatoria o proactiva de información (transparencia activa)
- Proyecto institucional de intervención documental (por ejemplo, organización de archivo histórico)
- Procesos de divulgación masiva, interoperabilidad o digitalización
- Desclasificación o apertura de fondos documentales, entre otros.

El alcance de la identificación dependerá del escenario identificado. En ese sentido, la entidad podrá aplicar los siguientes criterios:

- Identificación según el documento o conjunto documental objeto de intervención: Cuando la anonimización se origine en una solicitud concreta, la identificación deberá realizarse respecto del documento específico solicitado o del expediente relacionado. Cuando se trate de un proyecto institucional o intervención progresiva (por ejemplo, un fondo histórico o una serie documental), la identificación podrá realizarse por etapas, series, subseries o grupos documentales, conforme a una estrategia gradual definida por la entidad. No será exigible una revisión global e indiscriminada de todos los activos documentales, sino una identificación proporcional al alcance del proceso
- Verificación de presencia de datos personales: Una vez delimitado el documento o conjunto documental, la entidad deberá verificar si contiene datos personales, independientemente de su formato. Esta verificación deberá considerar:
 - Información estructurada
 - Información no estructurada
 - Elementos narrativos

- Información contenida en anexos, imágenes, audios o videos.

La identificación no podrá limitarse a la presencia de nombres propios; deberá evaluarse si el contenido permite vincular razonablemente el documento con una persona natural determinada o determinable.

- Clasificación del dato identificado: Los datos personales identificados deberán clasificarse según su naturaleza (públicos, semiprivados, privados o sensibles), con el fin de dimensionar el nivel de protección requerido. Cuando la entidad cuente con una matriz institucional de tipos de datos, esta deberá utilizarse como referencia técnica
- Identificación de identificadores directos e indirectos: La entidad deberá diferenciar entre: identificadores directos, que permiten la identificación inmediata, e identificadores indirectos o contextuales, que en combinación pueden permitir la reidentificación. En documentos no estructurados, la identificación deberá realizarse de manera contextual, evaluando la narrativa, la singularidad del hecho descrito y el entorno en el cual la información será divulgada.

Nota: El Registro de Activos de Información, el Índice de Información Clasificada y Reservada y los instrumentos de descripción documental deberán utilizarse como herramientas de apoyo cuando el escenario lo requiera, especialmente en procesos institucionales amplios o intervenciones progresivas. No obstante, su uso no sustituye el análisis concreto del documento objeto de intervención.

13.2 Determinación del régimen jurídico aplicable

La determinación del régimen jurídico aplicable constituye una fase decisiva dentro del proceso de anonimización. Su finalidad no es únicamente identificar la norma pertinente, sino definir si la información debe ser anonimizada, si está sometida a un régimen especial de reserva o si procede el acceso total, parcial o la negación.

Antes de aplicar cualquier técnica de intervención, la entidad deberá verificar si el documento identificado se encuentra sujeto exclusivamente al régimen general de transparencia y protección de datos, o si está cobijado por una legislación especial que imponga restricciones adicionales, tales como normas en materia de inteligencia y contrainteligencia, defensa y seguridad nacional, derechos humanos, justicia transicional, protección reforzada de niños, niñas y adolescentes o protección de víctimas. Se sugiere que la entidad tenga en cuenta los siguientes lineamientos:

- Verificación según fase del ciclo de vida documental: La entidad deberá determinar si el documento pertenece a: archivo de gestión (activo), archivo central, archivo histórico. Esta verificación es obligatoria, dado que el régimen de acceso y las restricciones pueden variar según la fase documental, especialmente en fondos históricos o procesos de desclasificación. En archivos históricos, el principio general es la publicidad responsable; sin embargo, ello no elimina la obligación de proteger datos personales, información sensible o sujetos de especial protección
- Identificación del escenario de divulgación: La entidad deberá establecer si la divulgación se produce en el marco de transparencia activa, transparencia pasiva, proyecto institucional de intervención documental, desclasificación de fondos, interoperabilidad o intercambio interinstitucional. El escenario condiciona el estándar de análisis jurídico
- Verificación de información clasificada o reservada: La entidad deberá determinar si el documento contiene información clasificada o reservada conforme al régimen general de acceso a la información pública. Cuando exista reserva legal o clasificación expresa, deberá establecerse si procede negación total, acceso parcial mediante anonimización, levantamiento de reserva por vencimiento de término
- Identificación de sujetos de especial protección: La entidad deberá verificar si el documento involucra: niños, niñas y adolescentes, víctimas de violencia o del conflicto armado, personas en situación de vulnerabilidad, personas cuyos datos sensibles puedan generar riesgo o discriminación. En estos casos, deberá aplicarse medidas reforzadas de análisis, aun cuando el documento tenga naturaleza pública
- Verificación de legislación especial aplicable: Cuando exista legislación especial, esta prevalecerá en lo que respecta a las reglas de acceso, reserva o protección reforzada. La anonimización no podrá utilizarse para desconocer una reserva legal expresa, ni para divulgar información cuya publicidad esté prohibida por norma especial. La entidad deberá analizar si el documento está sometido a un régimen jurídico especial que establezca reglas particulares de acceso, reserva o tratamiento, como, por ejemplo:
 - Normativa sobre inteligencia y contrainteligencia
 - Normativa de defensa y seguridad nacional
 - Régimen de protección de derechos humanos
 - Normativa sobre víctimas del conflicto armado
 - Legislación sectorial específica (salud, infancia, justicia, entre otras)
- Determinación del tipo de acceso procedente: Basado en el análisis anterior, la entidad deberá definir si el acceso será total, parcial mediante anonimización, negado por existencia de reserva o prohibición legal. El acceso parcial constituye un mecanismo de armonización

entre el derecho de acceso a la información y la protección de datos personales, siempre que no exista prohibición expresa de divulgación.

13.3 Intervención técnica sobre el documento de archivo

La intervención técnica sobre documentos de archivo constituye una fase operativa, mediante la cual se materializa la decisión jurídica de anonimizar. En esta etapa, la entidad debe aplicar criterios técnicos, archivísticos y de seguridad de la información para garantizar que la versión pública del documento no permita la identificación directa o indirecta de personas, sin afectar la integridad, autenticidad y valor probatorio del original. Dado que los documentos pueden encontrarse en múltiples soportes: físicos, digitales, audiovisuales o híbridos, la intervención debe adaptarse al formato específico, considerando los riesgos propios de cada uno y asegurando coherencia en todo el conjunto documental.

La anonimización de información no estructurada exige una intervención directa sobre el contenido. A diferencia de las bases de datos estructuradas, donde los campos pueden tratarse de manera sistemática, en los textos narrativos, imágenes, audios, videos y documentos físicos la información se encuentra integrada en el contexto, en la historia y en los detalles. Por ello, anonimizar no consiste únicamente en ocultar visualmente un dato. Consiste en impedir razonablemente que una persona pueda ser identificada de manera directa o indirecta a partir del contenido divulgado.

La regla general es que el documento original no se altera. El documento original conserva valor histórico, probatorio y archivístico. Alterarlo compromete la integridad documental y puede afectar la cadena de custodia, la autenticidad y la memoria institucional. La intervención debe realizarse siempre sobre una copia de trabajo, a partir de esta se genera una versión pública anonimizada, dejando trazabilidad del proceso aplicado. Sin esta separación entre original y versión intervenida, se compromete la seguridad jurídica y documental de la entidad.

13.4 Intervención en formato físico

La intervención en documentos de archivo en soporte físico exige un estándar reforzado de cuidado técnico y archivístico, dado que el soporte papel presenta riesgos específicos de recuperación de información. En este tipo de documentos, el tachado no elimina el dato; únicamente lo cubre. Bajo determinadas condiciones de iluminación, presión, escaneo o tratamiento digital posterior, el contenido ocultado puede hacerse visible total o parcialmente,

comprometiendo la efectividad de la anonimización. Por lo anterior, se sugiere adoptar los siguientes lineamientos:

- La entidad deberá abstenerse de intervenir el documento original y realizar toda actuación exclusivamente sobre una copia de trabajo previamente generada a partir del original. Esta regla es fundamental y tiene por finalidad preservar la integridad, autenticidad y valor probatorio del documento de archivo
- La intervención manual deberá efectuarse utilizando materiales que garanticen opacidad suficiente para impedir la lectura subyacente. Sin embargo, la copia intervenida no podrá considerarse automáticamente como versión pública. Con el fin de evitar la recuperación del contenido por relieve, transparencia o contraste digital, deberá generarse una segunda reproducción, mediante nueva fotocopia o digitalización controlada, que constituirá la versión pública definitiva. La copia intermedia intervenida no deberá circular y su eliminación deberá realizarse conforme a los protocolos de seguridad documental de la entidad
- Cuando una hoja contenga exclusivamente información sensible que carezca de valor contextual para la comprensión pública del documento, podrá retirarse de la copia de trabajo. No obstante, esta actuación deberá dejar constancia expresa mediante hoja de reemplazo que indique la supresión realizada y su fundamento jurídico, asegurando que la foliación y la estructura del expediente conserven coherencia archivística. La anonimización no puede generar vacíos documentales que no tengan explicación, ni alterar la unidad del expediente
- La revisión deberá extenderse no solo al cuerpo principal del documento, sino también a carátulas, encabezados, sellos, firmas manuscritas, anotaciones, índices, anexos, reversos de hoja y cualquier elemento accesorio que contenga datos personales. La anonimización parcial limitada al texto central resulta insuficiente cuando subsisten identificadores visibles en otros componentes del soporte físico
- Antes de autorizar la divulgación, la entidad deberá verificar que la versión pública no permita recuperación razonable del contenido eliminado y que el proceso haya sido ejecutado conforme a los estándares definidos. La inobservancia de estas condiciones compromete la eficacia de la anonimización y puede generar responsabilidad institucional.

13.5 Intervención en formato digital textual

La intervención en documentos digitales textuales (Word, PDF, correos y documentos electrónicos de archivo) exige un estándar técnico superior al aplicado en soporte físico, debido a que los riesgos de exposición no siempre son visibles. En archivos electrónicos pueden subsistir capas de información que no se ven a simple vista, tales como comentarios, control de cambios, historial de versiones, metadatos, propiedades del documento, capas OCR invisibles o texto

cubierto gráficamente sin eliminación real del contenido subyacente. En consecuencia, la anonimización en formato digital no podrá limitarse a cubrir visualmente información, sino que deberá garantizar la eliminación efectiva del dato en todas sus capas técnicas. Conforme a lo anterior la entidad deberá:

- Abstenerse de intervenir el archivo maestro o versión original del documento electrónico. Toda actuación deberá realizarse sobre un duplicado previamente generado, preservando intacto el documento original conforme a los principios de integridad, autenticidad y conservación archivística
- Durante la intervención, deberá verificarse que los datos personales hayan sido eliminados y no únicamente ocultados. En el caso de documentos en formato PDF, solo podrá utilizarse la función de redacción o censura permanente que suprima técnicamente el texto subyacente. Si el contenido intervenido puede copiarse, pegarse, seleccionarse o recuperarse mediante herramientas básicas de edición o extracción de texto, no podrá considerarse anonimizado
- Revisar y depurar comentarios, notas internas, marcas de revisión, versiones anteriores y cualquier rastro de edición que contenga datos personales. Los metadatos asociados al archivo, incluyendo autor, fecha de creación, historial de modificaciones, nombres de usuario y rutas de almacenamiento deberán ser eliminados antes de su divulgación, debido a que también pueden constituir datos personales o permitir inferencias indirectas
- Desde el punto de vista jurídico y de coherencia narrativa, cuando se opte por la sustitución de identificadores, esta deberá realizarse mediante etiquetas o códigos consistentes a lo largo del documento y sus anexos. La denominación asignada a un sujeto deberá mantenerse uniforme en todo el expediente, evitando inconsistencias que puedan generar confusión o facilitar procesos de reidentificación por contraste interno
- En casos de mayor riesgo, y especialmente cuando el documento sea objeto de publicación o difusión masiva, podrá requerirse el aplanamiento o rasterización del archivo como medida complementaria para impedir la recuperación técnica del contenido eliminado. Esta decisión deberá adoptarse atendiendo al nivel de riesgo contextual y al entorno de divulgación
- Antes de autorizar la publicación o entrega, la entidad deberá verificar técnicamente que la intervención ha sido efectiva y que no subsisten capas ocultas de información. La anonimización digital deficiente no solo compromete la protección de datos personales, sino que puede generar responsabilidad institucional por divulgación indebida.

13.6 Intervención en imágenes y fotografías

Las imágenes y fotografías constituyen uno de los formatos de mayor riesgo en materia de anonimización, dado que contienen tanto identificadores evidentes como elementos contextuales que pueden permitir la identificación directa o indirecta de una persona. La intervención sobre material gráfico no puede limitarse al rostro del titular, deberá considerar todos los componentes visuales que, individualmente o en combinación, permitan su singularización.

En una imagen pueden coexistir múltiples tipos de identificadores, tales como rasgos faciales, placas de vehículos, fachadas de viviendas, uniformes institucionales, distintivos laborales, tatuajes, cicatrices, marcas corporales, objetos personales, paisajes reconocibles o referencias geográficas específicas. Incluso cuando el rostro ha sido intervenido, el entorno puede proporcionar información suficiente para inferir la identidad del titular, especialmente en contextos territoriales reducidos o en hechos de amplia difusión pública. Considerando lo anterior, la entidad deberá:

- Realizar una evaluación integral del contenido visual antes de su divulgación, considerando no solo el elemento principal de la imagen sino también el fondo, los objetos accesorios, los rótulos visibles, las pantallas, los documentos capturados incidentalmente y cualquier otro componente que pueda permitir la reidentificación
- La intervención visual deberá ser técnicamente robusta. El pixelado, desenfoque o distorsión aplicada deberá impedir razonablemente la reconstrucción del elemento identificado. No se considerará suficiente una intervención superficial que permita revertir o mejorar digitalmente la imagen mediante herramientas de edición comunes. En casos de alto riesgo, podrá requerirse la cobertura total del área identificadora o la sustitución parcial de la imagen por versión recortada que conserve únicamente los elementos necesarios para la comprensión pública del documento
- Antes de su publicación o entrega, deberán eliminarse los metadatos asociados a la imagen. Los archivos digitales en muchas ocasiones incorporan información técnica como puede ser: coordenadas GPS, fecha y hora exacta de captura, modelo de dispositivo, usuario creador o rutas internas de almacenamiento, que puede constituir dato personal o facilitar procesos de geolocalización e inferencia contextual. La supresión de metadatos forma parte del proceso de anonimización y no puede considerarse una actuación opcional
- Cuando la imagen haga parte de un expediente más amplio, la intervención deberá ser coherente con la anonimización aplicada en los demás soportes documentales. Si una persona ha sido identificada mediante código o etiqueta en el documento textual, la

intervención visual deberá impedir cualquier elemento que permita asociar la imagen con la identidad original

- Antes de autorizar la divulgación, la entidad deberá verificar que la imagen intervenida no permita recuperación razonable del contenido eliminado ni identificación por contraste, ampliación o análisis digital básico. La anonimización deficiente en material gráfico puede resultar lesiva, dado el alcance masivo que pueden tener las imágenes en entornos digitales.

13.7 Intervención en formato audio

El tratamiento de archivos de audio presenta un nivel de complejidad particular, debido a que en este formato coexisten riesgos de naturaleza biométrica y riesgos derivados del contenido narrativo. La voz humana constituye un identificador biométrico susceptible de reconocimiento, incluso cuando no se menciona el nombre del titular. De igual manera, el relato contenido en el audio puede incorporar datos contextuales que permitan la singularización de la persona, aun cuando los identificadores directos hayan sido suprimidos. Conforme a lo anterior, se recomiendan los siguientes lineamientos:

- La anonimización de material sonoro no puede limitarse a la eliminación de nombres propios o referencias explícitas. La entidad deberá evaluar tanto la identidad vocal como el contenido del discurso, considerando la posibilidad de reidentificación por reconocimiento de voz, por referencias territoriales específicas, por descripción de hechos únicos o por mención de circunstancias fácilmente asociables a una persona determinada
- Cuando la identificación biométrica represente un riesgo, la intervención podrá requerir distorsión real de la voz mediante modificación de tono, frecuencia o timbre, utilizando herramientas que impidan razonablemente su reconocimiento. La simple reducción de volumen, la superposición de ruido superficial o la alteración mínima del audio no constituyen medidas suficientes cuando la voz conserva rasgos identificables
- La entidad deberá suprimir o editar fragmentos que contengan nombres propios, números de identificación, direcciones exactas, referencias laborales, familiares o cualquier elemento que permita la identificación directa. En los casos en que la referencia territorial o institucional sea excesivamente específica, podrá optarse por la generalización del dato, siempre que ello no afecte la comprensión sustantiva del contenido
- La evaluación deberá extenderse a situaciones en las que, aun sin mencionarse nombres, el relato describa un hecho único, mediático o fácilmente reconocible en un entorno determinado. La singularidad del evento puede permitir la identificación por inferencia, especialmente en comunidades pequeñas o en contextos ampliamente divulgados

- Cuando el audio haga parte de un expediente más amplio, la intervención deberá ser coherente con las etiquetas o códigos asignados en los documentos textuales o audiovisuales asociados, evitando inconsistencias que puedan facilitar la correlación de información
- Antes de autorizar la divulgación, la entidad deberá verificar que el archivo intervenido no permita recuperación del contenido eliminado ni reconocimiento razonable de la identidad vocal. La anonimización insuficiente en archivos de audio puede generar exposición masiva e irreversible, particularmente en entornos digitales donde la reproducción y distribución son inmediatas.

13.8 Intervención en formato video

El video constituye un formato de alta complejidad técnica en materia de anonimización, debido a que integra simultáneamente imagen, audio y contexto ambiental. A diferencia de otros soportes, en el video la información personal puede encontrarse distribuida en múltiples capas perceptibles y no perceptibles, lo que exige una intervención integral y coordinada. La anonimización en video no puede ser fragmentaria ni limitada a un solo elemento. La intervención deberá abordar de manera conjunta los componentes visuales, sonoros y contextuales, evaluando el riesgo de identificación directa, indirecta o por inferencia. Se sugiere a las entidades considerar los siguientes lineamientos:

- La entidad deberá intervenir los rostros mediante técnicas que impidan su reconocimiento razonable. Cuando exista movimiento, deberá aplicarse seguimiento dinámico para asegurar que la anonimización se mantenga durante toda la secuencia. La simple cobertura estática de un fotograma resulta insuficiente si el sujeto se desplaza o reaparece en otros ángulos. Asimismo, deberán evaluarse y, de ser necesario, intervenir otros elementos identificadores, tales como placas de vehículos, distintivos institucionales, uniformes personalizados, pantallas visibles, documentos capturados incidentalmente o señales particulares del entorno
- En el componente sonoro, deberá analizarse la necesidad de distorsionar la voz cuando esta constituya un identificador biométrico relevante. La supresión de nombres propios no elimina el riesgo si la voz conserva rasgos reconocibles o si el relato describe hechos fácilmente asociables a una persona determinada. En estos casos, podrá requerirse modificación técnica de tono y frecuencia, así como edición de fragmentos específicos
- El entorno ambiental también debe ser objeto de análisis. Elementos arquitectónicos singulares, referencias geográficas específicas, nombres visibles en fachadas, paisajes característicos o cualquier rasgo contextual que permita ubicar espacialmente el hecho

pueden facilitar la identificación, incluso cuando los sujetos hayan sido intervenidos. La evaluación deberá considerar el nivel de exposición pública del evento y la posibilidad de contraste con fuentes externas

- La intervención parcial en video puede ser insuficiente cuando subsisten elementos correlacionables en otras capas del mismo archivo o en documentos asociados. Por esta razón, el proceso deberá verificarse en su integridad antes de autorizar la divulgación, asegurando que no existan segmentos sin intervención, superposiciones accidentales o inconsistencias entre imagen y audio
- Cuando el video haga parte de un expediente más amplio, la anonimización deberá ser coherente con las etiquetas o códigos asignados en los documentos textuales, imágenes o audios relacionados, evitando contradicciones que permitan reconstruir la identidad por comparación interna
- Antes de su publicación o entrega, la entidad deberá verificar que el archivo final no permita recuperación razonable de la identidad mediante ampliación, mejora digital, análisis de fotogramas o reconocimiento básico de voz. Dada la facilidad de reproducción y circulación del video en entornos digitales, una anonimización deficiente puede generar afectaciones irreversibles a los derechos fundamentales de las personas involucradas.

13.9 Intervención en expedientes híbridos

Cuando un expediente combine soportes físicos, digitales y multimedia, la anonimización deberá aplicarse de manera transversal y coherente en todo el conjunto documental. La intervención no puede fragmentarse por tipo de formato, ni realizarse de manera aislada sobre cada soporte, ya que el riesgo de reidentificación aumenta cuando la información puede correlacionarse entre distintos componentes del mismo expediente. Se sugiere que:

- El expediente se entienda como una unidad documental integral. En consecuencia, la decisión técnica, adoptada respecto de la identificación de una persona, deberá mantenerse uniforme en todos los soportes asociados. Si un sujeto es sustituido por una etiqueta o código específico en el documento textual, por ejemplo, [TESTIGO A] o [FUNCIONARIO 1], esa misma denominación deberá utilizarse en las versiones digitalizadas, en audios, en videos, en transcripciones, en anexos gráficos y en cualquier otro elemento vinculado al expediente
- La inconsistencia en la denominación o la intervención parcial de los soportes puede permitir la reconstrucción de la identidad por contraste interno. Por ejemplo, la anonimización correcta en el documento escrito pierde eficacia si en el video asociado permanece un elemento visual identificable o si en el audio se conserva la identidad vocal

sin intervención. Del mismo modo, la supresión en el soporte físico resulta insuficiente, si la versión digital conserva metadatos o anexos sin tratamiento

- La entidad deberá realizar una revisión cruzada del expediente completo antes de autorizar su divulgación, verificando que la anonimización aplicada en cada soporte sea coherente, consistente y suficiente para impedir la identificación directa o indirecta. Esta revisión deberá considerar la posibilidad de correlación entre soportes y la eventual consulta simultánea de los distintos formatos por parte del usuario
- La anonimización en expedientes híbridos exige, por tanto, una visión sistémica y no fragmentaria. La unidad documental prevalece sobre el soporte individual, y la protección efectiva de los derechos fundamentales depende de la coherencia integral del proceso aplicado.

13.10 Intervención en elementos accesorios y capas técnicas no visibles

Existen elementos accesorios y capas técnicas no visibles que pueden contener datos personales o permitir la reidentificación indirecta, aun cuando el cuerpo del documento haya sido intervenido correctamente. Estos componentes forman parte del entorno documental y deben ser objeto de revisión obligatoria.

En soporte físico: En documentos en papel, la entidad deberá evaluar no solo el contenido textual interno, sino también los elementos externos o accesorios asociados al expediente, tales como: carátulas, etiquetas adhesivas, rótulos laterales, marcaciones manuscritas, códigos visibles en cajas o carpetas, índices impresos y cualquier referencia externa que permita identificar a una persona natural.

Cuando el documento vaya a ser consultado físicamente por terceros, deberá garantizarse que estos elementos no expongan datos personales durante la manipulación o visualización del expediente. La anonimización será incompleta si el contenido interno está intervenido, pero el entorno físico permite identificar al titular.

En soporte digital: En documentos electrónicos, la entidad deberá revisar y eliminar toda capa técnica no visible que pueda contener datos personales o información contextual relevante. Esto incluye, entre otros:

- Metadatos del archivo
- Propiedades del documento

- Historial de versiones
- Información incrustada en capas OCR
- Datos de autoría automática
- Rutas internas de almacenamiento
- Información técnica asociada al dispositivo de creación
- Archivos embebidos o vinculados.

La anonimización digital no será efectiva si el contenido visible ha sido intervenido, pero subsisten datos personales en las propiedades técnicas del archivo.

13.11 Verificación técnica final y validación previa a la divulgación

Concluida la intervención técnica sobre el documento o expediente, la entidad deberá realizar una verificación final antes de autorizar su publicación o entrega. Esta verificación constituye un control obligatorio destinado a confirmar que la anonimización aplicada es suficiente, coherente y efectiva frente al riesgo de identificación directa o indirecta.

La revisión final deberá efectuarse sobre la versión pública definitiva, no sobre la copia intermedia ni sobre el archivo maestro, y tendrá como finalidad comprobar que no subsisten datos personales visibles, ocultos o recuperables mediante medios razonables. Esta evaluación deberá:

- Abarcar tanto el contenido principal como anexos, metadatos, elementos contextuales, capas técnicas ocultas y cualquier otro componente asociado
- En el caso de expedientes híbridos, la verificación deberá realizarse de manera transversal, asegurando coherencia entre soportes físicos, digitales y audiovisuales, y descartando la posibilidad de reconstrucción de identidad por contraste entre ellos
- La entidad deberá confirmar que:
 - La intervención aplicada corresponde al régimen jurídico previamente determinado
 - No existen inconsistencias en la denominación de sujetos sustituidos
 - No es posible copiar, extraer, ampliar o mejorar técnicamente el contenido intervenido para recuperar información
 - La anonimización no afecta la comprensión sustancial del documento ni altera su integridad documental
 - La decisión de divulgar la versión pública deberá quedar formalmente validada por el área competente, dejando constancia interna del cumplimiento de las fases

metodológicas y del control técnico final. Esta validación forma parte de la trazabilidad del proceso y constituye garantía de seguridad jurídica institucional.

La omisión de la verificación final compromete la eficacia de la anonimización y puede generar responsabilidad por divulgación indebida de información protegida.

13.12 Medidas compensatorias cuando la anonimización no sea viable

Cuando la entidad no cuente con los recursos técnicos, tecnológicos, presupuestales o humanos suficientes para aplicar procesos de anonimización sobre documentos que contengan datos personales, deberá implementar medidas compensatorias orientadas a controlar, limitar y supervisar el acceso a la información, con el fin de reducir el riesgo de identificación de los titulares y proteger sus derechos fundamentales. Esto no sustituye definitivamente la anonimización, pero reduce el riesgo de identificación y permite cumplir el marco legal de forma proporcional.

Estas medidas constituyen mecanismos válidos de protección, en cumplimiento de los principios de seguridad, confidencialidad y circulación restringida establecidos en la Ley 1581 de 2012, así como del deber de acceso responsable a la información pública consagrado en la Ley 1712 de 2014. La implementación de estas medidas no exime a la entidad del deber de aplicar anonimización cuando esta sea técnica y operativamente viable, pero permite gestionar el riesgo de manera proporcional mientras se desarrollan o fortalecen las capacidades institucionales. Cuando no sea posible aplicar anonimización, la entidad deberá adoptar una o varias de las siguientes medidas:

- Consulta en entorno físico controlado: Permitir el acceso a los documentos únicamente en espacios designados por la entidad, bajo condiciones que garanticen el control y la supervisión de la información. La entidad deberá:
 - Habilitar una sala o espacio específico para la consulta documental
 - Supervisar la consulta por personal autorizado
 - Limitar el acceso simultáneo cuando sea necesario
 - Garantizar que los documentos no sean retirados del área autorizada.
- Restricción de dispositivos de captura y reproducción: Con el fin de evitar la reproducción o difusión no autorizada de la información, la entidad podrá restringir el ingreso o uso de

dispositivos que permitan copiar o capturar el contenido documental. Estas restricciones podrán incluir:

- Prohibición del uso de teléfonos móviles
 - Prohibición del uso de cámaras fotográficas o de video
 - Prohibición del uso de escáneres portátiles
 - Prohibición del uso de dispositivos de almacenamiento externo.
- Consulta en entornos digitales controlados: Cuando la consulta se realice en formato digital, la entidad deberá habilitar equipos o entornos tecnológicos que limiten la extracción, copia o transferencia de la información. Estas medidas podrán incluir:
 - Uso de equipos sin acceso a internet
 - Restricción de puertos USB y otros medios de almacenamiento externo
 - Restricción de funciones de copia, descarga o envío
 - Uso de perfiles de acceso con permisos limitados
 - Implementación de entornos de consulta supervisados.
 - Registro y trazabilidad del acceso: La entidad deberá mantener mecanismos que permitan identificar y registrar el acceso a los documentos, con el fin de garantizar la trazabilidad y el control institucional. El registro podrá incluir:
 - Identificación del usuario
 - Fecha y hora de acceso
 - Documentos consultados
 - Finalidad de la consulta, cuando corresponda.
 - Acceso supervisado: La entidad podrá disponer que la consulta de determinados documentos se realice bajo la supervisión directa de personal autorizado, especialmente cuando el contenido incluya datos personales sensibles o información de sujetos de especial protección constitucional. La supervisión tiene como finalidad prevenir la reproducción, uso o divulgación indebida de la información
 - Acuerdos de confidencialidad y de acceso: La entidad podrá requerir que los usuarios suscriban compromisos de confidencialidad y uso adecuado de la información, en los cuales se establezcan condiciones y responsabilidades relacionadas con el acceso, uso y divulgación de los documentos consultados. Estos compromisos tienen como finalidad reforzar el uso legítimo de la información y prevenir su utilización indebida. Este tipo de acuerdos de confidencialidad son útiles cuando el acceso es legítimo, pero existe riesgo de uso indebido

- Acreditación e identificación de usuarios que acceden a documentos históricos: El acceso a documentos históricos constituye una manifestación del derecho fundamental de acceso a la información pública, por lo que podrá ser ejercido por cualquier ciudadano, sin necesidad de acreditar interés particular o calidad específica, como investigador o académico. No obstante, con el fin de garantizar la seguridad de la información, la trazabilidad institucional y la protección de los derechos fundamentales de las personas cuyos datos puedan estar contenidos en los documentos, la entidad deberá implementar mecanismos de identificación y registro de los usuarios que acceden a la información. Este registro no constituye una restricción al derecho de acceso, sino una medida de seguridad orientada a prevenir el uso indebido de la información y garantizar el cumplimiento de las obligaciones legales en materia de protección de datos personales y gestión documental.

La entidad deberá aplicar estas medidas de manera proporcional, considerando el tipo de documento, la naturaleza de los datos contenidos, el riesgo de identificación, la finalidad del acceso, y las capacidades institucionales disponibles.

13.13 Mecanismos institucionales de acceso controlado cuando la anonimización no sea viable

El acceso a documentos que contienen datos personales debe garantizarse en armonía con los principios de acceso a la información pública, protección de datos personales y conservación documental. Si bien la anonimización constituye el mecanismo principal para proteger la identidad de los titulares cuando la información es divulgada o consultada, las entidades también deben implementar medidas institucionales orientadas a controlar, limitar y supervisar el acceso a la información, con el fin de reducir el riesgo de identificación indebida, prevenir la divulgación no autorizada y garantizar el uso legítimo de los documentos. Estas medidas resultan especialmente relevantes en fondos documentales que conservan su integridad archivística y valor histórico, y permiten a la entidad cumplir de manera proporcional y responsable con sus obligaciones legales, mientras se garantiza simultáneamente el derecho de acceso a la información pública y la protección de los derechos fundamentales de las personas.

- Acceso mediado a la información: cuando los documentos contengan datos personales y no sea técnica u operativamente viable aplicar procesos de anonimización, la entidad deberá privilegiar el acceso mediado a la información, evitando la entrega directa e íntegra del expediente, cuando ello no resulte necesario para satisfacer la finalidad legítima de la consulta. El acceso mediado consiste en proporcionar al usuario únicamente la información específica solicitada, sin entregar la totalidad del documento o expediente,

cuando este contenga datos personales cuya divulgación no sea necesaria. Para tal efecto, la entidad deberá:

- Analizar el alcance de la solicitud de acceso, con el fin de determinar la información estrictamente necesaria para satisfacer el requerimiento
- Extraer, transcribir o comunicar únicamente la información pertinente, omitiendo aquellos elementos que contengan datos personales no relevantes para la finalidad de la consulta
- Abstenerse de entregar el expediente completo cuando la finalidad del acceso pueda satisfacerse mediante la entrega parcial o mediada de la información
- Garantizar que el acceso mediado no altere el contenido original del documento, ni afecte su integridad o valor archivístico.

El acceso mediado permite garantizar el derecho de acceso a la información pública, reduciendo al mismo tiempo el riesgo de divulgación innecesaria de datos personales, en aplicación de los principios de finalidad, necesidad y circulación restringida, establecidos en la Ley 1581 de 2012.

- Restricción temporal o progresiva del acceso: La entidad podrá establecer condiciones de acceso temporal o progresivo respecto de documentos que contengan datos personales, cuando la divulgación irrestricta pueda afectar derechos fundamentales o generar riesgos de identificación indebida, especialmente en aquellos casos en que no sea viable aplicar anonimización inmediata. Estas medidas podrán incluir:
 - Restricción temporal del acceso a determinados documentos o series documentales, conforme a los criterios establecidos en las Tablas de Retención Documental (TRD), Tablas de Valoración Documental (TVD) u otros instrumentos archivísticos aplicables
 - Habilitación de acceso progresivo, condicionado a la implementación futura de procesos de anonimización o medidas técnicas que reduzcan el riesgo de identificación
 - Establecimiento de condiciones específicas de acceso, tales como consulta supervisada o acceso mediado, mientras se adoptan medidas definitivas de protección.

Estas restricciones deberán aplicarse de manera proporcional, razonable y temporal, y no podrán constituir una negación arbitraria del derecho de acceso a la información pública. Su finalidad es garantizar la protección de los derechos fundamentales de las personas, en armonía con el principio de acceso responsable a la información pública.

- Minimización del acceso a la información: La entidad deberá aplicar el principio de minimización del acceso, garantizando que los usuarios accedan únicamente a la información estrictamente necesaria para la finalidad legítima de la consulta, evitando la exposición innecesaria de datos personales contenidos en los documentos. En aplicación de este principio, la entidad deberá:
 - Limitar el acceso al documento, expediente o información directamente relacionada con la solicitud presentada
 - Abstenerse de proporcionar acceso a fondos documentales completos, expedientes integrales o conjuntos documentales amplios, cuando ello no sea necesario para satisfacer la finalidad de la consulta
 - Evaluar previamente el alcance del acceso solicitado, con el fin de reducir la exposición de datos personales no pertinentes
 - Implementar mecanismos de acceso parcial, acceso mediado o consulta controlada, cuando corresponda.

Esta regla permite reducir el riesgo de identificación indebida, proteger los derechos de los titulares de los datos personales y garantizar el cumplimiento de los principios de finalidad y circulación restringida, establecidos en la Ley 1581 de 2012.

Con el fin de facilitar la comprensión integral del proceso metodológico descrito en el presente capítulo y promover su aplicación coherente en las entidades públicas, se incorpora a continuación un instrumento de síntesis y verificación, que consolida las etapas del proceso de anonimización de información no estructurada.

Este instrumento no sustituye el desarrollo normativo contenido en los numerales precedentes, ni introduce obligaciones adicionales, sino que cumple una función orientadora para reforzar la trazabilidad, la proporcionalidad en la intervención y la adecuada articulación entre las fases jurídicas y técnicas del proceso.

Fase	Numeral	Propósito	Resultado esperado
Identificación	12.1	Determinar datos susceptibles de anonimización	Documento analizado
Régimen jurídico	12.2	Definir habilitación legal y límites	Concepto jurídico
Intervención técnica	12.3 – 12.10	Aplicar técnica según formato	Versión intervenida

Fase	Numeral	Propósito	Resultado esperado
Verificación final	12.11	Validar integridad y riesgo residual	Documento validado
Alternativas	12.12 – 12.13	Activar medidas compensatorias cuando aplique	Acceso controlado o decisión motivada

Fuente: elaboración propia.

14. Lineamientos para la selección de técnicas de anonimización en información no estructurada

La selección de la técnica de anonimización deberá realizarse atendiendo al tipo de soporte documental, al nivel de riesgo contextual identificado y al alcance de la divulgación prevista. No existe una técnica única aplicable a todos los formatos; por el contrario, la entidad deberá adoptar un enfoque diferenciado, y, cuando sea necesario, combinar varias modalidades de intervención para garantizar que la probabilidad de reidentificación mediante medios razonables sea baja. En dicho sentido, la entidad debe:

- En documentos textuales digitales tales como archivos en formato Word, PDF o correos electrónicos institucionales, la técnica recomendada será la sustitución uniforme de identificadores por etiquetas o códigos consistentes dentro del expediente. Esta técnica permite preservar la coherencia narrativa y la comprensión del contenido. No obstante, deberá complementarse obligatoriamente con la eliminación de metadatos, historial de versiones, comentarios y cualquier capa técnica oculta. Cuando el análisis contextual evidencie riesgo de singularización por combinación de atributos, deberá aplicarse adicionalmente la generalización de datos específicos, como fechas exactas, cargos únicos o ubicaciones precisas
- En documentos físicos en soporte papel, la técnica base consistirá en la supresión de datos personales sobre copia de trabajo, nunca sobre el documento original. Para garantizar que la información no pueda recuperarse por contraste o transparencia, deberá generarse una segunda reproducción que constituya la versión pública definitiva. La intervención deberá extenderse a carátulas, rótulos, etiquetas y cualquier elemento accesorio visible durante la consulta
- En imágenes y fotografías, la técnica principal será la distorsión visual robusta, mediante pixelado de alta densidad o cobertura opaca que impida la reconstrucción razonable del elemento identificador. Esta intervención deberá complementarse con la eliminación de

metadatos asociados al archivo y con la evaluación del entorno visual, dado que la identificación puede derivarse no solo del rostro sino de elementos contextuales como fachadas, placas, uniformes o señales geográficas

- En archivos de audio, cuando exista riesgo de reconocimiento biométrico, deberá aplicarse distorsión técnica de la voz mediante modificación de tono y frecuencia. Esta técnica deberá acompañarse de la supresión o edición de fragmentos que contengan referencias directas a nombres, ubicaciones exactas o circunstancias familiares. En casos donde el relato describa hechos únicos o ampliamente conocidos, podrá requerirse generalización narrativa adicional
- En material audiovisual, la anonimización deberá ser integral. El video exige intervención simultánea de imagen, sonido y entorno. Cuando exista movimiento del sujeto, deberá aplicarse seguimiento dinámico de los elementos identificadores. La distorsión de voz será obligatoria cuando el reconocimiento vocal represente un riesgo, y deberán eliminarse placas, pantallas visibles u otros identificadores ambientales. La intervención parcial se considerará insuficiente cuando subsistan elementos correlacionables en otras capas del mismo archivo
- En expedientes híbridos que integren soporte físico, digital y multimedia, la técnica recomendada será la codificación uniforme transversal, asegurando que la misma etiqueta se mantenga en todos los formatos asociados. La anonimización deberá aplicarse de manera coherente en todo el conjunto documental, evitando contradicciones o inconsistencias que permitan reconstrucción de identidad por contraste interno. En estos casos, la revisión cruzada integral antes de la divulgación será obligatoria
- Cuando la información no estructurada se encuentre integrada a sistemas de publicación dinámica o consulta pública, la entidad deberá aplicar técnicas adicionales como generalización de variables, supresión de campos sensibles y control de exportación masiva. La anonimización en entornos digitales de acceso amplio exige pruebas técnicas que verifiquen la imposibilidad de extracción o reconstrucción mediante medios automatizados
- La intensidad de la técnica aplicada deberá graduarse según el nivel de riesgo identificado. En escenarios de riesgo bajo, podrá resultar suficiente la supresión o sustitución básica acompañada de revisión técnica estándar. En escenarios de riesgo medio, será necesario combinar la técnica base con generalización adicional y eliminación rigurosa de metadatos, así como reforzar la revisión contextual. En escenarios de riesgo alto o de publicación masiva, deberán combinarse múltiples técnicas, pudiendo requerirse disociación irreversible, restricción de acceso o divulgación parcial debidamente motivada
- En todos los casos, la técnica seleccionada deberá ser coherente con el análisis de riesgo contextual previamente realizado y con el entorno de divulgación previsto. A mayor

amplitud de difusión, mayor deberá ser el estándar de intervención aplicado. La anonimización no podrá considerarse efectiva si se limita a ocultaciones superficiales que permitan recuperación técnica del contenido o reidentificación por correlación indirecta.

15. **Análisis de riesgo contextual y probabilidad de reidentificación**

La anonimización no puede considerarse como cumplida únicamente por la eliminación de identificadores directos e indirectos. La entidad deberá evaluar si, aun después de la intervención técnica, subsiste la posibilidad razonable de identificación directa o indirecta del titular mediante procesos de correlación, inferencia o vinculación con información disponible en el entorno. El análisis de riesgo contextual constituye una fase obligatoria del proceso de anonimización y deberá realizarse antes de la publicación o entrega de la versión pública. Su resultado deberá quedar documentado. La entidad deberá analizar, como mínimo, los siguientes factores:

- **Singularidad estructural:** Deberá evaluarse si el documento describe características únicas o altamente distintivas dentro de un grupo reducido que permitan individualizar razonablemente al titular. La singularidad puede derivarse del cargo desempeñado, la excepcionalidad del evento, el perfil profesional específico o condiciones personales poco frecuentes
- **Combinación de atributos:** Deberá analizarse si la combinación de variables tales como edad exacta, fecha precisa, municipio, cargo o circunstancias particulares reduce el universo de posibles titulares a una o pocas personas. Este análisis deberá considerar el tamaño del universo poblacional relevante y el contexto territorial
- **Vinculabilidad externa:** La entidad deberá verificar si la información puede correlacionarse con fuentes externas razonablemente accesibles, tales como registros públicos, noticias, boletines institucionales, redes sociales o bases abiertas de datos. La anonimización no podrá evaluarse de manera aislada del entorno informacional
- **Inferencia narrativa:** En documentos no estructurados, deberá evaluarse si el relato describe hechos únicos, mediáticos o fácilmente reconocibles que permitan identificar al titular aun sin datos directos
- **Entorno de divulgación:** La probabilidad de reidentificación deberá valorarse teniendo en cuenta el nivel de exposición previsto. A mayor amplitud de difusión, mayor exigencia en el estándar de anonimización
- **Valoración cualitativa del riesgo:** Con base en los criterios anteriores, la entidad deberá clasificar el riesgo de reidentificación como bajo, medio o alto:

- Se considerará bajo cuando la identificación requiera esfuerzos desproporcionados o acceso a información no razonablemente disponible
- Se considerará medio cuando exista posibilidad razonable de identificación mediante combinación de fuentes accesibles
- Se considerará alto cuando la identificación sea probable con información pública o fácilmente inferible
- Cuando el riesgo sea medio o alto, la entidad deberá reforzar la anonimización mediante generalización adicional, supresión contextual o restricción del acceso. Si el riesgo persiste de manera significativa, deberá optarse por acceso parcial reforzado o negación motivada.
- **Requisito de efectividad:** La anonimización se considerará adecuada únicamente cuando la probabilidad de reidentificación mediante medios razonables sea baja, atendiendo a la naturaleza de los datos, el tamaño del universo de referencia, el entorno territorial, la amplitud de divulgación y la disponibilidad de fuentes externas correlacionables. Se debe considerar que el análisis de riesgo contextual constituye un requisito indispensable del proceso de anonimización. Su omisión convierte la intervención en una supresión formal de datos, pero no en una anonimización efectiva.

16. Estructura institucional y responsabilidades

La estructura institucional encargada de los procesos de anonimización no es uniforme ni estática. Su configuración deberá definirse en función del escenario que active el proceso, del volumen documental involucrado, del nivel de riesgo y del tipo de divulgación prevista. La anonimización puede originarse, entre otros, en los siguientes escenarios:

- Respuesta a una solicitud concreta de acceso a la información (transparencia pasiva)
- Publicación proactiva o actualización de información institucional (transparencia activa)
- Proyectos de apertura o digitalización masiva de archivos históricos
- Procesos de interoperabilidad o intercambio interinstitucional
- Corrección posterior a divulgaciones indebidas o revisión extraordinaria.

Cada uno de estos escenarios exige una configuración organizacional proporcional y una aplicación coherente de los lineamientos metodológicos establecidos en el numeral 12 del presente documento. En todos los casos, la anonimización deberá desarrollarse conforme a una secuencia institucional que incluya, como mínimo: identificación del documento o expediente, determinación del régimen aplicable, identificación de datos susceptibles de anonimización,

evaluación de riesgo contextual, aplicación de la técnica correspondiente según formato, revisión jurídica y técnica, y registro de trazabilidad.

Configuración mínima para solicitudes puntuales: Cuando la anonimización derive de una solicitud individual de acceso, la entidad podrá conformar un esquema de intervención focalizado, en el que participen principalmente:

- El área productora del documento
- La Oficina Jurídica o quien ejerza el análisis del régimen aplicable
- Gestión Documental, cuando implique intervención sobre expediente físico
- Tecnologías de la Información, cuando la intervención requiera ajuste técnico específico

En estos casos, el equipo podrá ser reducido y actuar sobre un universo documental delimitado, siempre que se garantice revisión jurídica y técnica suficiente y que se documente cada etapa del proceso conforme a la secuencia metodológica institucional.

Configuración ampliada para proyectos institucionales o publicación masiva: Cuando la anonimización forme parte de una política de publicación activa, digitalización histórica, apertura masiva o rediseño de sistemas de consulta pública, la entidad deberá adoptar una configuración institucional ampliada. En estos escenarios deberán involucrarse, además de las áreas anteriores:

- Secretaría General o instancia coordinadora
- Oficina de Tecnologías de la Información con rol estructural
- Oficial de Protección de Datos
- Responsable de Seguridad de la Información
- Transparencia y Acceso a la Información
- Comunicaciones, cuando la divulgación tenga alcance público amplio
- Relatoría o dependencias especializadas, cuando aplique

En estos casos, la anonimización deja de ser una actuación puntual y se convierte en un proceso estructural que requiere planificación previa, aplicación ordenada de las fases metodológicas, trazabilidad y control continuo.

La entidad deberá aplicar un criterio de proporcionalidad en la conformación del equipo y en la intensidad del análisis. A mayor volumen documental, mayor exposición pública o mayor riesgo contextual, mayor deberá ser el nivel de articulación entre dependencias y control institucional. No resulta adecuado aplicar la misma estructura organizacional a un caso aislado que a un proyecto de apertura masiva de archivos históricos o a la implementación de un sistema de publicación automatizada.

Rol transversal de coordinación: Independientemente del escenario, la entidad deberá definir una instancia responsable de coordinar el proceso y asegurar coherencia en los criterios aplicados. Esta función no implica que dicha dependencia asuma toda la intervención técnica, sino

que garantice consistencia institucional, articulación entre dependencias y documentación integral de cada una de las etapas del proceso de anonimización.

Consideraciones tecnológicas: Las entidades públicas deben avanzar en la anonimización de información no estructurada de manera progresiva, proporcional y alineada con su nivel de madurez tecnológica, reconociendo que no todas parten del mismo nivel de capacidad institucional, infraestructura digital o integración sistémica. Desde una perspectiva de arquitectura empresarial, la anonimización no debe concebirse como una herramienta aislada ni como una intervención puramente operativa, sino como una capacidad transversal que interactúa con las capas de negocio, información, aplicaciones y tecnología de la entidad. En este sentido, se plantea un enfoque escalonado por niveles de madurez:

Entidades con baja madurez tecnológica: Aquellas que operan con infraestructura básica, escasa integración sistémica y procesos mayoritariamente manuales deberán estructurar la anonimización como un proceso controlado apoyado en protocolos formales, listas de verificación y revisión cruzada institucional. En este nivel, la gobernanza organizacional compensa la limitación tecnológica, garantizando que la intervención documental preserve la integridad del original, la coherencia narrativa y la protección efectiva de datos personales.

Entidades con gestión documental consolidada y madurez intermedia: En estos escenarios, la anonimización puede integrarse parcialmente a los sistemas de información institucionales, particularmente al gestor documental electrónico, permitiendo automatización asistida, aplicación estandarizada de reglas de sustitución y generación controlada de versiones públicas. La capacidad institucional se fortalece mediante interoperabilidad básica, trazabilidad automatizada y controles técnicos sobre metadatos y capas documentales.

Entidades con alta capacidad tecnológica: Aquellas que cuentan con arquitectura empresarial definida, interoperabilidad avanzada, gobierno de datos consolidado, integración entre aplicaciones misionales y sistemas de gestión documental, así como capacidades técnicas especializadas, deberían incorporar la anonimización como un componente estructural dentro de sus flujos de trabajo digitales. En este nivel de madurez, la anonimización deja de ser una acción reactiva o aislada y pasa a configurarse como una capacidad institucional integrada al ciclo de vida de la información, alineada con la arquitectura de información, la política de seguridad digital y el modelo de gobierno de datos de la entidad. Ello implica:

- Integración con el gestor documental institucional y sus reglas de retención
- Automatización controlada de identificación de datos personales
- Aplicación sistemática de reglas de reemplazo consistentes en todo el expediente
- Eliminación estructural de metadatos técnicos y ocultos
- Generación de versiones públicas auditables
- Registro automático de trazabilidad y bitácoras de intervención
- Validación jurídica previa y posterior integrada al flujo de aprobación

La anonimización, en este contexto, se convierte en una capacidad organizacional permanente, gestionada como servicio interno dentro del ecosistema digital institucional. Su diseño debe responder a principios de seguridad por diseño, privacidad por defecto y trazabilidad integral, asegurando que cada intervención sea verificable, replicable y auditable.

La incorporación de herramientas especializadas o desarrollos propios, cuando proceda, debe evaluarse bajo criterios funcionales y no comerciales, privilegiando aquellas soluciones que permitan:

- Eliminación efectiva y verificable de metadatos ocultos
- Consistencia automática de alias o sustituciones en todos los componentes del expediente
- Integración nativa o interoperabilidad segura con el gestor documental
- Registro de eventos y auditoría institucional
- Prevención de recuperación del texto oculto mediante extracción o análisis forense básico
- Adecuación al contexto lingüístico y jurídico nacional.

En este nivel de sofisticación, la anonimización no solo mejora la eficiencia operativa, sino que fortalece la gestión del riesgo institucional, especialmente en escenarios de alto volumen documental, publicación masiva o interoperabilidad interinstitucional. La automatización controlada, combinada con validación humana especializada, reduce la probabilidad de error, mejora la coherencia narrativa y asegura el cumplimiento simultáneo de los principios de transparencia, seguridad de la información y protección de datos personales.

No constituye anonimización segura la simple superposición de rectángulos gráficos, la exportación sin eliminación de capas editables, la sustitución manual sin control de consistencia o la permanencia del dato original en repositorios internos sin segmentación adecuada de acceso. Tales prácticas generan una falsa percepción de protección y exponen a la entidad a riesgos jurídicos, reputacionales y de reidentificación indirecta.

En consecuencia, el nivel de madurez tecnológica de la entidad deberá reflejarse no únicamente en la herramienta utilizada, sino en la integración coherente de la anonimización dentro de su arquitectura institucional, su modelo de gobierno de datos y su estrategia de transformación digital.

Participación del área misional o productora del documento: En todo proceso de anonimización deberá participar el área misional o dependencia que produjo el documento o que conoce el proceso de fondo del cual este se deriva. La intervención técnica o jurídica no puede realizarse de manera aislada del conocimiento sustantivo del contenido. La dependencia productora cumple un rol esencial en la identificación de:

- La finalidad institucional del documento
- La relevancia sustantiva de los datos contenidos

- El contexto en el que ocurrieron los hechos
- Los riesgos narrativos o de singularización
- La necesidad o no de conservar determinados elementos para garantizar comprensión pública.

La anonimización no puede vaciar el contenido sustancial ni alterar el sentido del documento. En consecuencia, el área misional deberá validar que la intervención aplicada preserve la coherencia narrativa y la finalidad institucional, sin comprometer la protección de los derechos fundamentales involucrados. Su participación es obligatoria tanto en solicitudes puntuales como en proyectos estructurales de publicación masiva, con intensidad proporcional al alcance del proceso.

Con el fin de facilitar la aplicación de los lineamientos anteriores y asegurar coherencia institucional en la asignación de responsabilidades, se presenta la siguiente matriz de referencia, la cual articula la secuencia metodológica del proceso con las dependencias que intervienen en cada fase.

Secuencia metodológica y responsabilidades institucionales en el proceso de anonimización de información no estructurada

Fase	Etapla metodológica	Alcance por tipo de información no estructurada	Dependencia líder	Dependencias que intervienen	Producto o evidencia
1	Identificación del documento, archivo o soporte	Papel, digital, imagen, audio, video, híbrido	Área productora	Gestión Documental	Registro de activación
2	Determinación del régimen aplicable	Todos los soportes	Oficina Jurídica	Área productora	Concepto jurídico
3	Identificación de datos susceptibles de anonimización	Todos los soportes	Gestión Documental	Área productora	Listado preliminar
4	Evaluación de riesgo contextual y narrativo	Todos los soportes	Jurídica	Área productora	Acta de análisis
5	Generación de copia de trabajo (cuando aplique)	Digital, imagen, audio, video	Gestión Documental	TI	Copia controlada
6	Aplicación de técnica específica según formato	Según soporte (redacción, distorsión,	Dependencia designada	TI	Versión intervenida

Fase	Etapas metodológicas	Alcance por tipo de información no estructurada	Dependencia líder	Dependencias que intervienen	Producto o evidencia
		edición, sustitución, etc.)			
7	Eliminación de metadatos y verificación técnica (si aplica)	Digital, imagen, audio, video	TI	Gestión Documental	Registro técnico
8	Revisión cruzada jurídica y técnica	Todos los soportes	Jurídica	Gestión Documental	Validación formal
9	Evaluación de riesgo residual	Todos los soportes	Jurídica	Área productora	Concepto final
10	Generación de versión pública	Todos los soportes	Gestión Documental	TI	Documento o archivo final
11	Registro de trazabilidad institucional	Todos los soportes	Gestión Documental	—	Bitácora institucional
12	Publicación o entrega	Todos los soportes	Dependencia competente	—	Evidencia de divulgación

Fuente: elaboración propia.

17. BIBLIOGRAFÍA

Normativa colombiana

Constitución Política

Colombia. Asamblea Nacional Constituyente. (1991). Constitución Política de Colombia. Bogotá D.C.: Imprenta Nacional. [Artículos 15, 74].

Leyes

- Colombia. Congreso de la República. (1972, 15 de enero). Ley 13 de 1972, sobre voto secreto. Diario Oficial n.º 33.752.
- Colombia. Congreso de la República. (1981, 18 de febrero). Ley 23 de 1981, por la cual se dictan normas en materia de ética médica. Diario Oficial n.º 35.711. [Artículos 33 y ss.: Reserva de la Historia Clínica].
- Colombia. Congreso de la República. (1989, 18 de agosto). Ley 80 de 1989, por la cual se crea el Archivo General de la Nación. Diario Oficial n.º 38.853.
- Colombia. Congreso de la República. (1996). Ley 256 de 1996, por la cual se dictan normas sobre competencia desleal. [Artículo 16: Secretos empresariales].
- Colombia. Congreso de la República. (2000, 14 de julio). Ley 594 de 2000, Ley General de Archivos. Diario Oficial n.º 44.084.
- Colombia. Congreso de la República. (2006, 8 de noviembre). Ley 1098 de 2006, Código de la Infancia y la Adolescencia. Diario Oficial n.º 46.446. [Artículo 47].
- Colombia. Congreso de la República. (2008, 31 de diciembre). Ley 1266 de 2008, por la cual se dictan disposiciones generales del habeas data. Diario Oficial n.º 47.219. [Literal g) y h), artículo 3].
- Colombia. Congreso de la República. (2011, 10 de junio). Ley 1448 de 2011, Ley de Víctimas y Restitución de Tierras. Diario Oficial n.º 48.096.
- Colombia. Congreso de la República. (2011, 18 de enero). Ley 1437 de 2011, Código de Procedimiento Administrativo y de lo Contencioso Administrativo. Diario Oficial n.º 47.956. [Artículo 24].
- Colombia. Congreso de la República. (2012, 17 de octubre). Ley 1581 de 2012, por la cual se dictan disposiciones generales para la protección de datos personales. Diario Oficial n.º 48.587. [Artículos 4, 5, 6 y 10].

- Colombia. Congreso de la República. (2013, 17 de abril). Ley 1621 de 2013, normas sobre actividades de inteligencia y contrainteligencia. Diario Oficial n.º 48.764. [Artículo 33].
- Colombia. Congreso de la República. (2013, 6 de marzo). Ley 1618 de 2013, disposiciones para el pleno ejercicio de los derechos de las personas con discapacidad. Diario Oficial n.º 48.717.
- Colombia. Congreso de la República. (2014, 6 de marzo). Ley 1712 de 2014, Ley de Transparencia y del Derecho de Acceso a la Información Pública Nacional. Diario Oficial n.º 49.084. [Artículos 2, 18, 19 y 21].
- Colombia. Congreso de la República. (2023). Ley 2335 de 2023, disposiciones sobre las estadísticas oficiales en el país. Diario Oficial.
- Colombia. Congreso de la República. (1986). Código Electoral. [Artículo 51].
- Colombia. Estatuto Tributario. [Artículo 583: Reserva de la información tributaria].

Decretos

- Colombia. Presidencia de la República. (2014, 21 de marzo). Decreto 857 de 2014, por el cual se reglamentan artículos de la Ley 1621 de 2013. Diario Oficial n.º 49.102.
- Colombia. Presidencia de la República. (2015, 26 de mayo). Decreto 1070 de 2015, Decreto Único Reglamentario del Sector Administrativo de Defensa. Diario Oficial n.º 49.523.
- Colombia. Presidencia de la República. (2015, 26 de mayo). Decreto 1074 de 2015, Decreto Único Reglamentario del Sector Comercio, Industria y Turismo. Diario Oficial n.º 49.523. [Artículo 2.2.2.25.1.3].
- Colombia. Presidencia de la República. (2015, 26 de mayo). Decreto 1080 de 2015, Decreto Único Reglamentario del Sector Cultura. Diario Oficial n.º 49.523. [Artículos 2.8.2.1.1 y ss.; artículo 2.8.5.2.1].
- Colombia. Presidencia de la República. (2015, 26 de mayo). Decreto 1081 de 2015, Decreto Único Reglamentario de la Presidencia de la República. Diario Oficial n.º 49.523. [Artículo 2.1.1.1.1 y ss.; artículo 31: Anonimización de información mixta]. [Compilación del Decreto 103 de 2015].
- Colombia. Presidencia de la República. (2017). Decreto 2149 de 2017, por el cual se crea el Sistema Nacional de Depuración de Datos y Archivos de Inteligencia y Contrainteligencia. Diario Oficial.
- Colombia. Presidencia de la República. (2025). Decreto 1400 de 2025, relacionado con la desclasificación de archivos del DAS. Diario Oficial.

Acuerdos del Archivo General de la Nación

- Archivo General de la Nación Jorge Palacios Preciado. (2024, 29 de febrero). Acuerdo 001 de 2024, por el cual se establece el Acuerdo Único de la Función Archivística. Bogotá D.C.: AGN. https://normativa.archivogeneral.gov.co/wp-content/uploads/2024/04/2024-02_29_AcuerdoAGN-FIRMADO.pdf

Normativa regional e internacional aplicable

- Comunidad Andina. (2000). Decisión 486: Régimen Común sobre Propiedad Industrial. [Secretos comerciales].

Jurisprudencia- Corte Constitucional de Colombia

- Corte Constitucional de Colombia. (2002, 5 de septiembre). Sentencia T-729 de 2002. M.P. Eduardo Montealegre Lynett.
- Corte Constitucional de Colombia. (2003, 4 de marzo). Sentencia C-185 de 2003. M.P. Eduardo Montealegre Lynnet. Fundamento jurídico 9.
- Corte Constitucional de Colombia. (2004, 8 de marzo). Sentencia T-216 de 2004. M.P. Eduardo Montealegre Lynett. Fundamento jurídico 21.
- Corte Constitucional de Colombia. (2005). Sentencia T-718 de 2005.
- Corte Constitucional de Colombia. (2008). Sentencia C-1011 de 2008.
- Corte Constitucional de Colombia. (2010). Sentencia T-656 de 2010.
- Corte Constitucional de Colombia. (2011, 6 de octubre). Sentencia C-748 de 2011. M.P. Jorge Ignacio Pretelt Chaljub. Fundamentos jurídicos 2.4.3.2 y 2.6.3.
- Corte Constitucional de Colombia. (2013, 6 de diciembre). Sentencia T-926 de 2013. M.P. Mauricio González Cuervo.
- Corte Constitucional de Colombia. (2014). Sentencia T-020 de 2014.

Jurisdicción Especial para la Paz (JEP)

- Jurisdicción Especial para la Paz. (s.f.). Auto 001. Medidas cautelares sobre archivos del DAS.
- Jurisdicción Especial para la Paz. (s.f.). Auto 073.
- Jurisdicción Especial para la Paz. (s.f.). Auto 182.
- Jurisdicción Especial para la Paz. (s.f.). Auto 778. Protocolos de acceso para el Sistema Integral de Verdad.

Documentos institucionales

Archivo General de la Nación (AGN)

- Archivo General de la Nación Jorge Palacios Preciado. (2020). Guía de anonimización de datos estructurados - Conceptos generales y propuesta metodológica. Bogotá D.C.: AGN.
- Archivo General de la Nación Jorge Palacios Preciado, & Centro Nacional de Memoria Histórica. (2017). Protocolo de gestión documental de los archivos referidos a las graves y manifiestas violaciones a los Derechos Humanos e infracciones al Derecho Internacional Humanitario, ocurridas con ocasión del conflicto armado interno. Bogotá D.C.: AGN – CNMH.

https://www.archivogeneral.gov.co/sites/default/files/Estructura_Web/5_Consulte/Recursos/Publicacionees/2022/protocolo-gd-ddhh.pdf

Consejo Superior de la Judicatura

- Consejo Superior de la Judicatura. (2024). Colección Guías Pedagógicas: Protección de datos personales, anonimización y seguridad de la información. Bogotá D.C.: Rama Judicial – Consejo Superior de la Judicatura. ISBN: 978-958-5570-45-0. Elaborado por Consorcio DEPIN 006 (InvesCor SAS y Deproyectos SAS, Contrato 130 de 2023). Autoras: Heidí Elieth Balanta y Susan Simoneth Suárez Gutiérrez. Disponible en:

<https://www.ramajudicial.gov.co/web/publicaciones/publicaciones-2019>

Centro Nacional de Memoria Histórica (CNMH)

- Centro Nacional de Memoria Histórica. (2015). Política Pública de Archivos de Derechos Humanos, Memoria Histórica y Conflicto Armado. Bogotá D.C.: CNMH. <https://centrodememoriahistorica.gov.co/descargas/politica-publica-archivo-ddhh.pdf>
- Centro Nacional de Memoria Histórica. (s.f.). Guía para la Anonimización de Datos e Información No Estructurada: Estándares y Lineamientos Técnicos. Bogotá D.C.: CNMH.

Superintendencia de Industria y Comercio (SIC)

- Superintendencia de Industria y Comercio. (2018). Concepto n.º 7075 de 2018. [Clasificación de imagen fotográfica y video]. Bogotá D.C.: SIC.
- Superintendencia de Industria y Comercio. (2014). Concepto n.º 14-268304. [Clasificación de datos educativos y académicos]. Bogotá D.C.: SIC.
- Superintendencia de Industria y Comercio. (2024). Circular Externa 002 de 2024, lineamientos sobre el tratamiento de datos personales en sistemas de inteligencia artificial. Bogotá D.C.: SIC.

Ministerio de Tecnologías de la Información y las Comunicaciones (MinTIC)

- Ministerio de Tecnologías de la Información y las Comunicaciones. (2016). Resolución 24109 de 2016. [Datos de contacto corporativo de servidores públicos]. Bogotá D.C.: MinTIC.

Obras académicas y literatura especializada

Libros y capítulos de libro

- López Coba, A. D. Habeas data y función archivística. En Documentos, mensajes de datos. Manuscrito inédito, Archivo General de la Nación, Bogotá D.C.

Artículos científicos y estudios técnicos

- Article 29 Working Party. (2014). Opinion 05/2014 on Anonymisation Techniques (WP216). Bruselas: Comisión Europea. https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf
- Lison, P., Pilán, I., Sánchez, D., Batet, M., & Øvrelid, L. (2022). The Text Anonymization Benchmark (TAB): A Dedicated Corpus and Evaluation Framework for Text Anonymization. Computational Linguistics, MIT Press. <https://direct.mit.edu/coli/article/48/4/1053/112770>
- Rocher, L., Hendrickx, J., & de Montjoye, Y. A. (2019). Estimating the success of re-identifications in incomplete datasets using generative models. Nature Communications. <https://www.nature.com/articles/s41467-019-10933-3>
- Todt, J., Hanisch, S., & Strufe, T. (2024). Fantômas: Understanding Face Anonymization Reversibility. Proceedings on Privacy Enhancing Technologies (PoPETs), 2024(4), 24–43. DOI: 10.56553/popets-2024-0105. <https://arxiv.org/abs/2210.10651>

- Staab, R., Vero, M., Balunović, M., & Vechev, M. (2024). Beyond Memorization: Violating Privacy Via Inference with Large Language Models. ICLR 2024. <https://arxiv.org/abs/2310.07298>
- Ishihara, S. (2023). Training Data Extraction From Pre-trained Language Models: A Survey. En: Proceedings of the 3rd Workshop on Trustworthy Natural Language Processing (TrustNLP 2023), pp. 260–275. Toronto: Association for Computational Linguistics. DOI: 10.18653/v1/2023.trustnlp-1.23. <https://aclanthology.org/2023.trustnlp-1.23/>
- Wiepert, D., et al. (2024). Reidentification of Participants in Shared Clinical Data Sets: Experimental Study. JMIR AI. <https://ai.jmir.org/2024/1/e52054>

Guías y documentos de autoridades de protección de datos

- Agencia Española de Protección de Datos (AEPD). (s.f.). Guía de Anonimización. Madrid: AEPD. <https://www.aepd.es>
- Agencia Española de Protección de Datos (AEPD), & Personal Data Protection Commission (PDPC) de Singapur. (2022). Guía sobre los conceptos de anonimización y seudonimización. Madrid / Singapur: AEPD – PDPC.
- Autoridade Nacional de Proteção de Dados (ANPD) de Brasil. (2024). Estudo Preliminar: Anonimização e Pseudonimização para a Proteção de Dados Pessoais. Brasília: ANPD. <https://www.gov.br/anpd/pt-br/assuntos/noticias/anpd-abre-consulta-a-sociedade-sobre-o-guia-de-anonimizacao-e-pseudonimizacao>

Estándares y principios internacionales

- Organización de Naciones Unidas (ONU). (1997). Principios de Joinet. Conjunto actualizado de principios para la protección y la promoción de los derechos humanos mediante la lucha contra la impunidad (E/CN.4/Sub.2/1997/20/Rev.1). Nueva York: ONU.
- Open Society Justice Initiative. (2013). Principios Globales sobre Seguridad Nacional y el Derecho a la Información (Principios de Tshwane). Nueva York: Open Society Foundations.
- ARTICLE 19. (1996). Principios de Johannesburg sobre Seguridad Nacional, Libertad de Expresión y Acceso a la Información. Londres: ARTICLE 19. <https://www.article19.org/resources/johannesburg-principles-national-security-freedom-expression-access-information/>
- Organización para la Cooperación y el Desarrollo Económicos (OCDE). (2013). Directrices de la OCDE sobre la Protección de la Privacidad y los Flujos Transfronterizos de Datos Personales (Revisión 2013). París: OCDE. <https://www.oecd.org/sti/ieconomy/privacy-guidelines.htm>

- Organización de las Naciones Unidas (ONU), Asamblea General. (2013). Resolución 68/167: El derecho a la privacidad en la era digital. Nueva York: ONU. <https://undocs.org/A/RES/68/167>
- Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos (OHCHR). (2014). El derecho a la privacidad en la era digital. Informe A/HRC/27/37. Ginebra: ONU. <https://undocs.org/A/HRC/27/37>
- Ochoa, S., Rasmussen, J., Robson, C., & Salib, M. (2007). Reidentification of Individuals in Chicago's Homicide Database: A Technical and Legal Study. Cambridge: Massachusetts Institute of Technology (MIT). <http://web.mit.edu/sem083/www/assignments/reidentification.html>
- Data Protection Commission (DPC) de Irlanda. (2019). Guidance on Anonymisation and Pseudonymisation. Dublín: DPC. <https://www.dataprotection.ie/sites/default/files/uploads/2019-06/190614%20Anonymisation%20and%20Pseudonymisation.pdf>

Fuentes digitales y recursos en línea

- Adobe Inc. (s.f.). ¿Qué es OCR o reconocimiento óptico de caracteres? Adobe Acrobat Resources. <https://www.adobe.com/es/acrobat/resources/ocr.html>

Versión 1.0

25 de febrero de 2026



Archivo General de la Nación



Conoce todas nuestras publicaciones



<https://www.archivogeneral.gov.co/recursos>

Síguenos en nuestras redes sociales



@CanalAGNColombia



@ArchivoGeneral



Archivo General de la Nación