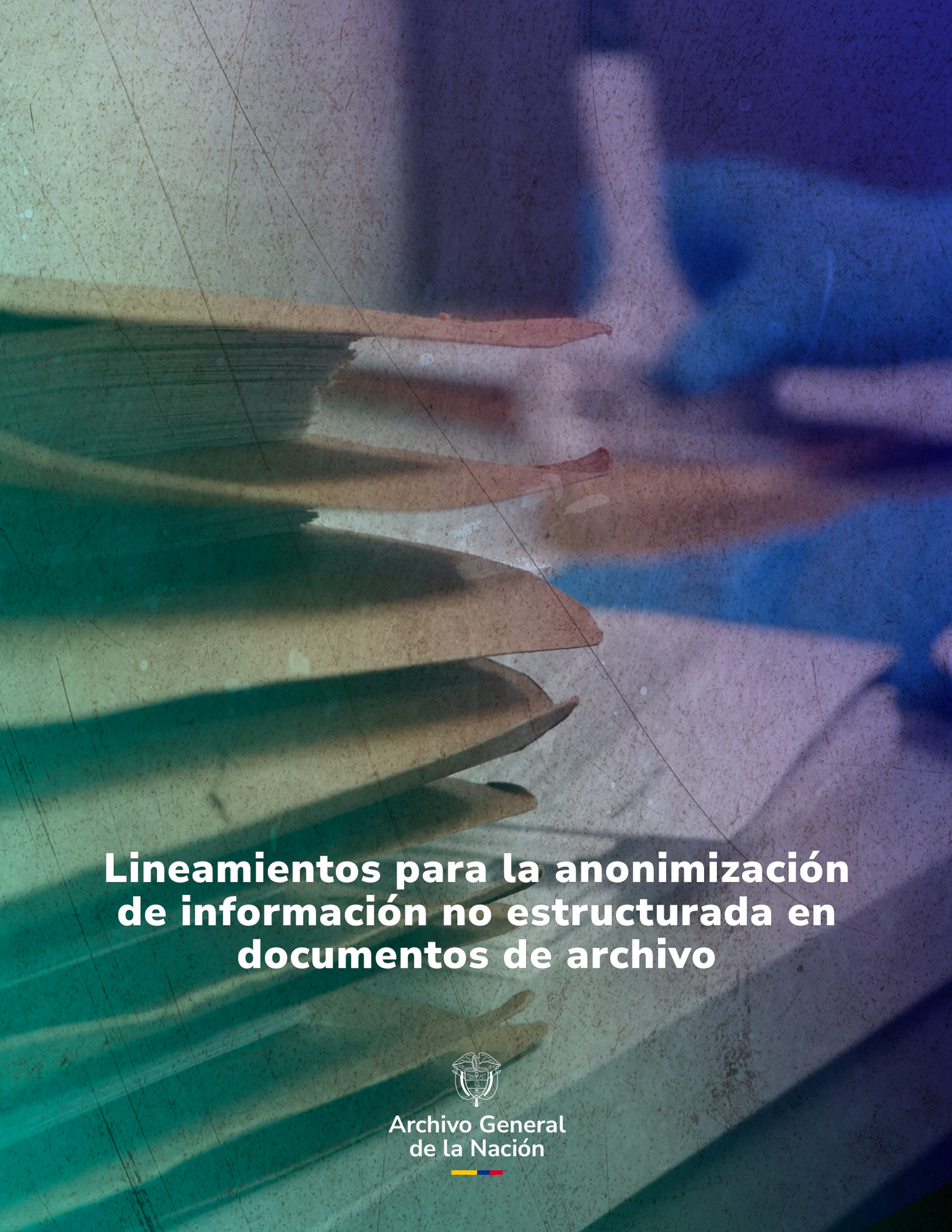


Lineamientos para la anonimización de información no estructurada en documentos de archivo

Archivo General
de la Nación

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

ÍNDICE DE TABLAS 5

OBJETIVO 5

ALCANCE 6

DESTINATARIOS 6

PRESENTACIÓN 6

INTERPRETACIÓN Y APLICACIÓN 8

¿CUÁNDO Y POR QUÉ ANONIMIZAR? 10

PRIMERA PARTE 12

Marco conceptual y fundamentos de la anonimización 12

CONCEPTOS FUNDAMENTALES PARA LA ANONIMIZACIÓN 13

PROTECCIÓN DE DATOS PERSONALES Y LA FUNCIÓN ARCHIVÍSTICA 17

Principios del régimen de protección de datos personales y su relación con la función archivística 19

SUPUESTOS EN LOS QUE PROCEDE ANALIZAR LA ANONIMIZACIÓN 22

BUENAS PRÁCTICAS TÉCNICAS INTERNACIONALES Y NACIONALES DE REFERENCIA 25

MARCO LEGAL DE LA ANONIMIZACIÓN EN DOCUMENTOS DE ARCHIVO 29

Ley 1581 de 2012. Régimen general de la protección de datos personales en Colombia 29

Ley 1712 de 2014. Régimen general de transparencia y acceso a la información pública 30

Ley 594 de 2000 - Ley General de Archivos 31

Acuerdo 001 de 2024 del Archivo General de la Nación *31*

Legislación complementaria 31

LA INFORMACIÓN NO ESTRUCTURADA EN LOS DOCUMENTOS DE ARCHIVO 34

Nuevas formas de producción de información no estructurada 35

Formatos no estructurados que pueden contener datos personales o información sensible 35

Datos específicos dentro de formatos no estructurados 37

Información no estructurada que se utiliza en las entidades públicas 38

TIPOS DE DATOS PERSONALES SUJETOS A ANONIMIZACIÓN: IDENTIFICADORES DIRECTOS E INDIRECTOS 40

Categorías de datos personales según su nivel de protección 40

Identificadores directos e indirectos 41

SUJETOS, ESPACIOS Y CONTEXTOS OBJETO DE PROTECCIÓN MEDIANTE PROCESOS DE ANONIMIZACIÓN. 47

Categoría de sujetos protegidos 47
Espacios, lugares y hechos objeto de protección 49

TÉCNICAS DE ANONIMIZACIÓN PARA INFORMACIÓN NO ESTRUCTURADA 52

Supresión o eliminación directa 54
Sustitución por etiquetas, codificación y tokenización semántica 56
Seudonimización controlada 58
Criterios para distinguir la anonimización de la seudonimización 59
Generalización 60
Disociación irreversible 61
Distorsión técnica en formatos audiovisuales 62
Enmascaramiento de caracteres 64
Resumen o abstracción 65
Agregación estadística 65
Combinación de técnicas y criterios de selección 67
Técnicas insuficientes o no admisibles por sí solas 68

RIESGOS DE REIDENTIFICACIÓN 69

Evaluación y registro del riesgo residual 70
Modalidades técnicas de riesgo de reidentificación 70
Impacto de la inteligencia artificial sobre las modalidades de riesgo 71
Test de reidentificación 73
Métricas cuantitativas de referencia para evaluar la anonimización 74
Las métricas de datos estructurados como referente conceptual 76
Balance riesgo-utilidad: el principio de proporcionalidad aplicado a la anonimización 76
Caso aplicado 78
Decisión institucional 80
Consecuencias de una reidentificación 80

SEGUNDA PARTE 82

Lineamientos específicos para la anonimización y protección de datos personales en la gestión documental 82

ANONIMIZACIÓN SEGÚN EL CICLO DE VIDA DOCUMENTAL 83

Consideración sobre niveles de intervención 83
Lineamiento de anonimización en archivos de gestión 85
Lineamiento de anonimización en archivo central 86
Lineamiento de anonimización en archivo histórico 88

LINEAMIENTOS METODOLÓGICOS EN EL PROCESO DE ANONIMIZACIÓN 90

Articulación con instrumentos archivísticos 90
Pasos para la articulación de los regímenes aplicables 91
Identificación de información susceptible de anonimización 96
Determinación del régimen jurídico aplicable 97
Intervención técnica sobre el documento de archivo 100
Verificación técnica final y validación previa a la divulgación 100
Prueba de no recuperación 101
Revisión cruzada entre soportes 102
Registro formal de trazabilidad de la intervención 102
Eliminación segura de datos en procesos de anonimización 102
Cuadro resumen del proceso metodológico 104

INTERVENCIÓN TÉCNICA SEGÚN SOPORTE DOCUMENTAL 104

Reglas generales 104
Intervención en formato físico 106
Intervención en formato digital textual 107
Intervención en imágenes y fotografías 108
Intervención en formato audio 110
Intervención en formato video 111
Intervención en expedientes híbridos 112
Intervención en elementos accesorios y capas técnicas no visibles 113

SELECCIÓN DE TÉCNICAS Y ANÁLISIS DE RIESGO CONTEXTUAL 114

Lineamientos para la selección de técnicas 114
Análisis de riesgo contextual aplicado a la selección de técnicas 115
Matriz de decisión: soporte, nivel de riesgo y técnicas recomendadas 117

MEDIDAS ALTERNATIVAS Y MECANISMOS DE ACCESO CONTROLADO 118

Medidas compensatorias cuando la anonimización no sea viable 118
Mecanismos institucionales de acceso controlado 120

ESTRUCTURA INSTITUCIONAL Y RESPONSABILIDADES 122

Configuración institucional según el escenario 122
Configuración mínima para solicitudes puntuales 123
Configuración ampliada para proyectos institucionales o publicación masiva 123
Rol transversal de coordinación 124
Consideraciones tecnológicas 124
Participación del área misional o productora del documento 126
Secuencia metodológica y responsabilidades institucionales en el proceso de anonimización de información no estructurada 127

BIBLIOGRAFÍA 129

Índice de tablas

Tabla 1. Principios del habeas data y su correlación con los principios archivísticos 19

Tabla 2. Buenas prácticas técnicas internacionales y nacionales de referencia 25

Tabla 3. Regímenes especiales, jurisprudencia constitucional y otros instrumentos internacionales 32

Tabla 4. Categorías de datos personales según su nivel de protección 42

Tabla 5. Técnicas insuficientes o no admisibles por sí solas 68

Tabla 6. Tres vías de reidentificación: singularización, vinculabilidad e inferencia 70

Tabla 7. Métricas cuantitativas de referencia para evaluar la anonimización 74

Tabla 8. Pruebas de reidentificación aplicadas al caso del informe trimestral 78

Tabla 9. Cuadro resumen del proceso metodológico de anonimización 104

Tabla 10. Matriz de decisión: soporte, nivel de riesgo y técnicas recomendadas 117

Tabla 11. Secuencia metodológica y responsabilidades institucionales 127

Objetivo

Orientar a los sujetos obligados, en el marco del ejercicio de la función archivística y la gestión documental, en la anonimización y seudonimización de información no estructurada contenida en documentos de archivo, con el fin de proteger a las personas, mitigar riesgos de reidentificación y revictimización, y contribuir a la articulación entre la función archivística, el régimen de protección de datos personales[1], y el derecho de acceso a la información pública, garantizando la transparencia, la legalidad del tratamiento y la calidad de la información publicada.

1 Para efectos de los presentes lineamientos, se utilizarán indistintamente los términos autodeterminación informativa, Habeas Data y protección de datos personales, conforme a la jurisprudencia de la Corte Constitucional. Al respecto, la Sentencia C-748 de 2011 (M.P. Jorge Ignacio Pretelt Chaljub), mediante la cual la Sala Plena ejerció el control de constitucionalidad de la Ley 1581 de 2012, señaló: "El título del proyecto de ley Por el cual se dictan disposiciones generales para la protección de datos personales indica que su objetivo principal es regular de manera general las garantías del derecho fundamental a la protección de los datos personales, derecho que en la jurisprudencia constitucional ha sido con frecuencia denominado derecho al habeas data y en algunas oportunidades derecho a la autodeterminación informática o informativa".

Alcance

El presente documento orienta los procesos de identificación, evaluación y aplicación de técnicas de anonimización y seudonimización sobre información no estructurada contenida en documentos de archivo, en el marco del ejercicio de la función archivística y la gestión documental por parte de los sujetos obligados, cuando dichos documentos contengan información cuya divulgación pueda afectar derechos fundamentales, la seguridad del Estado o los derechos de terceros, y sean objeto de descripción, acceso, publicación, divulgación o entrega a terceros, ya sea de manera proactiva o en virtud de solicitudes de acceso a la información pública. La aplicación de los presentes lineamientos se entiende sin perjuicio de las competencias de la Superintendencia de Industria y Comercio (SIC) como autoridad nacional de protección de datos personales, del Ministerio Público en materia de acceso a la información pública y de las autoridades sectoriales con competencias específicas sobre archivos sometidos a regímenes especiales.

Destinatarios

Los presentes lineamientos aplican a la administración pública en sus diferentes niveles: nacional, departamental, distrital y municipal, así como las entidades privadas que cumplen funciones públicas y demás organismos regulados por la Ley 594 de 2000.

Para efectos de los presentes lineamientos, se hace referencia al término de sujetos obligados, en concordancia con el ámbito de aplicación de la Ley 594 de 2000, Ley General de Archivos, para referir a las entidades públicas o privadas que cumplen funciones públicas, que por efecto incluye a: servidores públicos, trabajadores oficiales, personas naturales o jurídicas de derecho privado, que por mandato legal cumplen una función pública.

Presentación

La información en poder de las entidades públicas se rige por el derecho de acceso a la información pública consagrado en el artículo 74 de la Constitución Política, desarrollado en los artículos 2 y 3 de la Ley 1712 de 2014 y reconocido como uno de los principios de la función archivística en el literal f) del artículo 4 de la Ley 594 de 2000. Bajo este principio, el acceso es la regla y la restricción es la excepción. Los presentes lineamientos parten de ese punto de partida: no establecen un deber general de anonimizar los documentos de archivo de las entidades públicas, ni invierten la lógica del acceso. Su finalidad es ofrecer orientaciones archivísticas, metodológicas y de articulación

normativa para los escenarios específicos en que la divulgación íntegra de un documento puede afectar derechos fundamentales o intereses constitucionalmente protegidos, supuestos en los cuales corresponde a la entidad ponderar y, cuando proceda, intervenir el documento sin sacrificar el acceso al contenido públicamente relevante. 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.

La anonimización es una de las técnicas que permite armonizar el acceso a la información pública con la protección de los derechos de quienes aparecen en los documentos, cuando del análisis del caso concreto se desprende que la divulgación íntegra, puede afectar tales derechos o intereses constitucionalmente protegidos. No constituye una regla general aplicable a todo documento de archivo que contenga datos personales, ni un paso obligado en la divulgación de información pública. Su aplicación es residual y supone un ejercicio previo de ponderación contextualizada y jurídicamente fundamentado, que considere el régimen aplicable, el tipo de documento, el soporte, el contenido, los sujetos involucrados y el uso previsto de la información. Cuando del análisis se concluye que la divulgación no afecta derechos o intereses protegidos, el documento se difunde íntegro, sin intervención técnica. Cuando se acredita una afectación, la anonimización ofrece un camino para abrir el archivo sin desproteger a quienes en él aparecen.
El documento se estructura en dos partes. La primera desarrolla el marco conceptual y los fundamentos de la anonimización: define los conceptos esenciales; examina la relación entre la protección de datos personales y la función archivística; identifica los supuestos en que las entidades públicas tienen el deber de anonimizar; establece los principios orientadores del proceso; precisa el marco legal aplicable; caracteriza la información no estructurada en los documentos de archivo; delimita los tipos de datos sujetos a anonimización y los sujetos objeto de protección, y analiza las técnicas disponibles y los riesgos de reidentificación.

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.

En consecuencia, y como principio rector de la metodología aquí expuesta, se establece que la anonimización no autoriza la alteración, supresión o modificación del documento original que integra el expediente. Toda intervención deberá realizarse únicamente sobre reproducciones, copias de trabajo o versiones destinadas al acceso público, preservando el original íntegro como garantía de autenticidad e integridad, en concordancia con el artículo 4 de la Ley 594 de 2000.

Interpretación y aplicación

Los presentes lineamientos constituyen un documento orientador, elaborado por el Archivo General de la Nación Jorge Palacios Preciado, en ejercicio de sus funciones como ente coordinador y orientador 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. Cada entidad deberá adaptarlos a su naturaleza jurídica, funciones misionales, nivel de riesgo, capacidades técnicas y al contexto documental específico, evaluando en cada caso si la divulgación de un documento determinado afecta o no derechos fundamentales o intereses constitucionalmente protegidos. 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.

Sin perjuicio del alcance definido en el presente documento, orientado a la anonimización de información no estructurada contenida en documentos de archivo, estos lineamientos podrán articularse con modelos institucionales de gobierno de datos, arquitectura de información y gestión de sistemas de información, en aquellos escenarios en los que la información no estructurada tenga su origen, transformación o consumo en entornos digitales.

En este sentido, la anonimización podrá entenderse como una capacidad transversal que, si bien se materializa sobre documentos de archivo, puede requerir su integración con procesos, servicios y componentes tecnológicos asociados a la gestión de datos, con el fin de garantizar coherencia, trazabilidad y efectividad en su aplicación.

Para efectos de facilitar su aplicación, las entidades podrán considerar la anonimización dentro de un esquema analítico que identifique los distintos niveles en los que la información es generada, transformada y utilizada, tales como:

- el nivel de generación de información, asociado a sistemas, procesos o interacciones digitales
- el nivel de información no estructurada, donde se encuentran los contenidos objeto de análisis
- el nivel documental, en el que la información se consolida como documento de archivo
- el nivel de acceso y uso, en el que se materializa la divulgación o consulta.

Este enfoque no implica la intervención sobre dichos niveles, sino que permite mejorar la comprensión del contexto en el que se produce la información contenida en los documentos, fortaleciendo la identificación de riesgos y la aplicación de técnicas de anonimización.

Para la anonimización de datos estructurados, microdatos, registros administrativos en formato tabular y operaciones estadísticas, los presentes lineamientos remiten a la Guía para la anonimización de datos estructurados publicada por el Departamento Administrativo Nacional de Estadística (DANE, 2024), elaborada en el marco del Sistema Estadístico Nacional conforme a la Ley 2335 de 2023. Las entidades que apliquen los presentes lineamientos deberán consultar dicha guía, siempre que el documento o conjunto documental objeto de intervención contenga componentes estructurados, o cuando el proceso institucional de anonimización abarque, además de la información no estructurada, microdatos o bases de datos sujetas al régimen estadístico.

Para la anonimización de información no estructurada contenida en archivos referidos a graves y manifiestas violaciones a los derechos humanos, infracciones al DIH ocurridas con ocasión del conflicto armado interno, y archivos de memoria histórica en general, los presentes lineamientos se aplican de forma articulada con la Guía para la anonimización de datos e información no estructurada: estándares y lineamientos técnicos (CNMH-AGN, 2022) y con el Protocolo de gestión documental de los archivos referidos a las graves y manifiestas violaciones a los derechos humanos e infracciones al DIH ocurridas con ocasión del conflicto armado interno (Resolución conjunta AGN-CNMH 031 de 2017, versión II de 2022, de obligatoria adopción para los sujetos obligados conforme a la Ley 594 de 2000). Cuando exista coexistencia entre estos instrumentos, prevalecerá

la disposición que ofrezca mayor protección a los derechos del titular y a la finalidad de verdad, justicia, reparación y no repetición.

¿Cuándo y por qué anonimizar?

La sola presencia de datos personales en un documento no activa la anonimización. La anonimización procede únicamente cuando, del análisis del documento concreto y de su contexto de divulgación, se desprende que la difusión íntegra puede afectar derechos fundamentales o intereses constitucionalmente protegidos. Una vez identificado un supuesto que justifica la intervención, la anonimización ofrece un camino técnico y jurídico para abrir el documento sin desproteger a quienes en él aparecen. La anonimización es definida como sigue:

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].

En los supuestos en que procede, la anonimización hace posible ejercer de forma simultánea el derecho de acceso a la información pública y el derecho fundamental a la protección de datos personales, Habeas Data o autodeterminación informativa. Permite que la información circule sin exponer innecesariamente a quienes aparecen en los documentos cuando esa exposición puede afectar sus derechos. Bien aplicada, contribuye a la confianza ciudadana en las instituciones, en la medida en que el ciudadano sabe que el Estado abre sus archivos sin instrumentalizar su vida privada cuando ello no es necesario para el ejercicio del derecho de acceso.
La información no estructurada contenida en documentos de archivo (textos narrativos, imágenes, registros de audio y video, expedientes híbridos) plantea retos de anonimización cualitativamente distintos a los de los datos estructurados. En ella, los datos personales no están organizados en campos predefinidos sino dispersos en el contenido, entretejidos con el contexto y, con frecuencia, interdependientes entre sí. Esto exige criterios técnicos, jurídicos y archivísticos específicos, que son precisamente el objeto de los presentes lineamientos.

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.

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

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 estos.

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

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 (AGN, Acuerdo 001 de 2024)
- **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" (DANE, 2024)
- **Archivo:** 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 gestión administrativa, la información, la investigación y la cultura (AGN, Acuerdo 001 de 2024)
- **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:** Aquellos 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 (DANE, 2024)
- **Datos no estructurados:** Información 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 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:** 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, 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 y AEPD, 2022)

- **Digitalización:** Técnica que permite la reproducción de información que se encuentra guardada de manera analógica (papel, video, casettes, cinta, película, microfilm y otros) en una que solo puede leerse o interpretarse por sistemas de información (AGN, Acuerdo 001 de 2024)
- **Documento electrónico de archivo:** Registro de información generada, producida, 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): 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)[3]
- **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, códices, prensa, entre otros (AGN, Acuerdo 001 de 2024)

3 Definición propia, 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*, que traduce reconocimiento óptico de caracteres): tecnología que permite convertir imágenes de texto (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 (AGN, Acuerdo 001 de 2024)
- **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 y 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 (TRD), 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 (AGN, Acuerdo 001 de 2024)
- **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 (DANE, 2024)

- **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 (AGN, Acuerdo 001 de 2024)
- **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).

Protección de datos personales y la función archivística

Cuando se hace referencia a documentos, específicamente a información, ha de considerarse que esta contiene datos, incluso 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 el artículo 15 de la Constitución Política de Colombia:

> *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.*

La jurisprudencia de la Corte Constitucional[4], relacionada con *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*

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

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]cualquier 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, Sentencia C-748, 6 de octubre del 2011).

Y concluye la citada Corporación Judicial:

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 (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 (Sentencia C-748, 6 de octubre del 2011).

Por lo tanto, los datos personales deben ser tratados por la gestión archivística de manera que se garantice la privacidad y la protección de los datos de las personas. Además, tal información deberá ser organizada y clasificada según las disposiciones archivísticas aplicables a la preservación de los documentos, velando en que la información que pueda comprometer a una persona en particular o afectarla de alguna manera, conforme al régimen de protección de datos personales.

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

Los principios del régimen de protección de datos personales no operan de manera aislada ni en sustitución de los principios del régimen de transparencia y acceso a la información pública ni de los principios de la función archivística. Los tres conjuntos de principios convergen en la gestión de los documentos de archivo y se aplican simultáneamente: el principio de máxima publicidad establece la regla general de acceso a la información pública; los principios archivísticos garantizan la integridad, autenticidad y conservación del documento, y los principios del *Habeas Data* orientan la protección de los datos personales que el documento pueda contener. La tabla comparativa que se presenta a continuación se concentra en la relación entre el *Habeas Data* y la función archivística, sin que ello implique subordinación del régimen de transparencia, cuya operación queda recogida transversalmente a lo largo de los presentes lineamientos.

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:

Tabla 1. Principios del Habeas Data y su correlación con los principios archivísticos

(PRINCIPIOS) HABEAS DATA[5] Ley 1581 de 2012. Artículo 4°. Principios para el tratamiento de datos personales. Corte Constitucional. Sentencia C-748 del 2011		**FUNCIÓN ARCHIVÍSTICA. Ley 594 de 2000. Artículo 4. Principios generales. Corte Constitucional. Sentencia T-656 del 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 16. Obligaciones de los funcionarios a cuyo cargo estén los archivos de las entidades públicas

5 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 DATA[5] Ley 1581 de 2012. Artículo 4°. Principios para el tratamiento de datos personales. Corte Constitucional. Sentencia C-748 del 2011		FUNCIÓN ARCHIVÍSTICA. Ley 594 de 2000. Artículo 4. Principios generales. Corte Constitucional. Sentencia T-656 del 2010
(iv)	Principio de integridad en el manejo de datos	Principio de orden original. Ley 594 de 2000. Artículo 11. Los principios de procedencia y orden original están referidos en el artículo 11 de dicha ley
(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
(ix)	Principio de temporalidad, caducidad[6] o derecho al olvido (Supresión de datos)	Retención documental. Valoración documental. Disposición final de los documentos, selección, eliminación y 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, del libro Documentos, mensajes de datos. Arturo Daniel López Coba. (Sin publicar).

6 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 inadmisible la conservación indefinida de datos personales que revelen información negativa una vez hayan desaparecido las causas que justificaron su acopio y administración".

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.

> *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 información7.*
>
> *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 prestaciones (Sentencia T-926, 6 de diciembre del 2013)8.*

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 circulación (Sentencia T-216, 8 de marzo del 2004)9.*

---

7 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.

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

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

Supuestos en los que procede analizar la anonimización

La anonimización es una técnica cuyo análisis y eventual aplicación procede únicamente en contextos específicos en los que el ordenamiento jurídico, o las condiciones del caso concreto, hacen necesario ponderar la apertura íntegra del documento frente a la posible afectación de derechos fundamentales o intereses constitucionalmente protegidos. En todos los demás casos, el documento se difunde íntegro, conforme al principio de máxima publicidad. Los supuestos que se describen a continuación son los contextos más frecuentes en los que esa ponderación se hace necesaria en la práctica institucional. Su enunciación es ilustrativa y no taxativa: la procedencia del análisis depende de las circunstancias del documento, los sujetos involucrados y el régimen aplicable, y debe resolverse caso a caso. La conclusión del análisis puede ser anonimizar, aplicar otra técnica de protección, restringir el acceso por un tiempo determinado, o difundir el documento sin intervención.

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 cuya divulgación íntegra puede afectar derechos fundamentales o estar sujeta a clasificación o reserva, la entidad debe valorar las distintas vías que ofrece el ordenamiento: la entrega íntegra cuando proceda, la denegación motivada por reserva o clasificación conforme a los artículos 18 y 19 de la Ley 1712 de 2014, la entrega parcial mediante versión pública (artículo 21 de la Ley 1712 de 2014) o la anonimización del contenido protegido. La selección de la respuesta dependerá del análisis de cada caso.

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 y máxima publicidad.

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 NNA, los testigos y las fuentes en procesos judiciales, así como

la información relacionada con actividades de inteligencia, contrainteligencia, seguridad y defensa nacional.

Tratamiento con fines estadísticos, científicos o históricos

La Ley 1581 de 2012 prohíbe, como regla general, el tratamiento de datos sensibles aquellos que pueden afectar la intimidad o generar discriminación, como los referidos a la salud, el origen racial, la orientación sexual o las creencias religiosas, políticas o filosóficas. Sin embargo, su artículo 6, literal e, establece una excepción: cuando el tratamiento tenga una finalidad histórica, estadística o científica, los datos sensibles pueden ser tratados sin autorización del titular, siempre que se adopten las medidas conducentes a la supresión de su identidad. Esto significa que, en los procesos archivísticos con fines históricos, estadísticos o científicos, la anonimización es la condición legal que habilita el tratamiento.

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.

Seguridad y defensa nacional

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. Este supuesto se distingue del de información clasificada y reservada en que no opera sobre la categoría jurídica del documento, sino sobre el riesgo concreto que su divulgación representaría en un contexto operacional específico.

Pruebas tecnológicas, desarrollo de software, uso de Inteligencia Artificial

En aplicación de los principios de privacidad por diseño y privacidad por defecto, las entidades públicas que utilicen información institucional en ambientes de prueba, procesos de desarrollo de software, ejercicios de analítica interna, capacitaciones, pilotos tecnológicos o entrenamiento de sistemas de inteligencia artificial deben evaluar, en primer lugar, si el uso de datos personales reales cuenta con un habilitante legal válido (autorización del titular, ejercicio de funciones legales conforme al literal a, artículo 10 de la Ley 1581 de 2012, u otro supuesto del artículo 6 cuando

aplique). Cuando no exista habilitante legal aplicable, o cuando exista, pero la finalidad técnica no requiera identificar al titular, se recomienda adoptar la anonimización previa de la información.

Información clasificada y reservada

La Ley 1712 de 2014 establece dos categorías de información exceptuada del principio de máxima publicidad: la información pública clasificada, cuyo acceso puede ser rechazado cuando su divulgación cause un daño a derechos de personas naturales o jurídicas; y la información pública 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.

Proceso de desclasificación de la información o levantamiento de la reserva

El levantamiento de la reserva o la desclasificación de un documento no equivale automáticamente a la publicación íntegra de la información. Ambas figuras operan de manera distinta: la reserva está sujeta a plazos máximos conforme al artículo 22 de la Ley 1712 de 2014, mientras que la clasificación persiste mientras subsistan las razones que la motivaron conforme al parágrafo del artículo 18 de la misma ley. La anonimización no procede mientras la reserva legal esté vigente: durante ese período el acceso se rige por la regla de denegación total, conforme al numeral 12.4. La anonimización opera únicamente después del levantamiento, como mecanismo para preservar los datos personales que conservan protección indefinida bajo el artículo 18 (intimidad, derecho a la vida o seguridad, secretos comerciales), aun cuando la reserva del artículo 19 haya vencido. En ambos casos, 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 o que el titular haya otorgado su consentimiento. Por lo tanto, una vez levantada la reserva o desclasificado el documento, se debe generar la identificación de la información personal que continúa protegida, y finalmente la anonimización como condición para la creación de la versión pública del documento. Los supuestos descritos evidencian que la decisión sobre anonimizar no responde a un criterio único, ni se activa por la sola presencia de datos personales en un documento. Procede en contextos

normativos, institucionales y operacionales específicos, en los que es necesario un análisis previo que considere el régimen aplicable, los derechos en juego, la finalidad de la divulgación y las técnicas disponibles. Cuando del análisis se concluye que la divulgación íntegra no afecta derechos ni intereses protegidos, el documento se difunde sin intervención. Cuando se acredita una afectación, la anonimización es una de las técnicas disponibles para abrir el documento sin desproteger a quienes en él aparecen.

Buenas prácticas técnicas internacionales y nacionales de referencia

El diseño y aplicación del proceso de anonimización puede beneficiarse de la consulta de buenas prácticas y técnicas reconocidas. Estas referencias no son parte del ordenamiento jurídico colombiano, ni tienen carácter vinculante, pero ofrecen criterios técnicos útiles para fortalecer el rigor metodológico del proceso. A continuación, se relacionan algunas buenas prácticas:

Tabla 2. Buenas prácticas técnicas internacionales y nacionales de referencia

Categoría	Referente	Descripción	Acceso
Estándar técnico	**ISO/IEC 20889:2018** *Privacy enhancing data de-identification terminology and classification of techniques* (revisada y confirmada por ISO en 2024; versión nacional ANSI: INCITS/ISO/IEC 20889:2018(R2024))	Estándar internacional sobre terminología y clasificación de técnicas de desidentificación	iso.org/standard/69373.html
Estándar técnico	**NIST IR 8053 (2015)** De-Identification of Personal Information	Reporte interno del NIST que ofrece un panorama general de técnicas de desidentificación y reidentificación. Antecedente conceptual de SP 800-188	csrc.nist.gov/pubs/ir/8053/final
Estándar técnico	**NIST SP 800-188 (septiembre de 2023)** De-Identifying Government Datasets: Techniques and Governance	Guía especializada para entidades gubernamentales sobre desidentificación, gobernanza del proceso y técnicas formales como la privacidad diferencial	csrc.nist.gov/pubs/sp/800/188/final

Categoría	Referente	Descripción	Acceso
Guía / dictamen autoridad PDP	**Dictamen 05/2014 del WP29** sobre técnicas de anonimización	Referente conceptual aún citado: criterios de singularización, vinculabilidad e inferencia. **Observación:** en proceso de actualización; el EDPB tiene en su Programa de Trabajo 2026-2027 la adopción de Guidelines específicas sobre anonimización	ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf
Guía / dictamen autoridad PDP	EDPB, Guidelines 01/2025 on Pseudonymisation (adoptadas el 16 de enero de 2025)	Lineamientos europeos vigentes sobre seudonimización; introducen el concepto de *pseudonymisation domain* y precisan la condición de dato personal de la información seudonimizada	edpb.europa.eu/our-work-tools/documents/public-consultations/2025/guidelines-012025-pseudonymisation_en
Guía / dictamen autoridad PDP	EDPB, Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models (17 de diciembre de 2024)	Establece el test caso a caso para considerar anónimo un modelo de IA y pautas sobre extracción de datos personales del modelo	edpb.europa.eu/our-work-tools/our-documents/opinion-board-art-64/opinion-282024-certain-data-protection-aspects_en
Guía / dictamen autoridad PDP	AEPD, Orientaciones y garantías en los procedimientos de anonimización de datos personales (2016)	Guía estructural sobre fases del proceso, equipo, EIPD y gestión del riesgo residual de reidentificación	aepd.es/guias/guia-orientaciones-procedimientos-anonimizacion.pdf
Guía / dictamen autoridad PDP	**AEPD, Guía básica de anonimización** (2022, traducción de la PDPC de Singapur)	Guía didáctica orientada a *pymes* y *startups*, con herramienta práctica gratuita en Excel	aepd.es/documento/guia-basica-anonimizacion.pdf
Guía / dictamen autoridad PDP	PDPC Singapur, Guide to Basic Anonymisation (versión actualizada, julio de 2024)	Original de la guía traducida por la AEPD; referente práctico internacional	pdpc.gov.sg
Guía / dictamen autoridad PDP	ICO Reino Unido, Anonymisation, pseudonymisation and privacy enhancing technologies guidance (28 de marzo de 2025)	Introduce el espectro de identificabilidad y la prueba del intruso motivado. **Observación:** bajo revisión desde la	ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/data-sharing/anonymisation/

Categoría	Referente	Descripción	Acceso
		entrada en vigor del *Data (Use and Access) Act* (19 de junio de 2025)	
Guía / dictamen autoridad PDP	CNIL Francia, Guide de la sécurité des données personnelles (edición 2024)	Incluye fichas sobre anonimización y seudonimización dentro del marco de seguridad RGPD	cnil.fr/sites/cnil/files/2024-03/cnil_guide_securite_personnelle_2024.pdf
Guía / dictamen autoridad PDP	ENISA, Pseudonymisation Techniques and Best Practices (2019)	Análisis técnico inicial de escenarios y técnicas de seudonimización	enisa.europa.eu/publications/pseudonymisation-techniques-and-best-practices
Política pública internacional	**OECD** (2023) Emerging Privacy-Enhancing Technologies: Current Regulatory and Policy Approaches, Digital Economy Papers No. 351	Taxonomía de PETs y panorama regulatorio comparado	oecd.org/en/publications/emerging-privacy-enhancing-technologies_bf121be4-en.html
Política pública internacional	**OECD (2025)** Sharing trustworthy AI models with privacy-enhancing technologies, OECD AI Papers No. 38	Aplicación de PETs (privacidad diferencial, TEE, cifrado homomórfico, federated learning) en modelos de IA	oecd.org/en/publications/sharing-trustworthy-ai-models-with-privacy-enhancing-technologies_a266160b-en.html
Principio DD. HH. / acceso a información	Principios de Johannesburgo (1996)	Principios sobre seguridad nacional, libertad de expresión y acceso a la información	article19.org
Principio DD. HH. / acceso a información	Principios de Tshwane (2013)	Principios globales sobre seguridad nacional y el derecho a la información	opensocietyfoundations.org
Principio DD. HH. / acceso a información	**Principios Joinet/Orentlicher** (ONU, 2005) Documento E/CN.4/2005/102/Add.1	Conjunto de principios actualizados para la protección y promoción de los derechos humanos mediante la lucha contra la impunidad	ohchr.org
Principio archivístico	Declaración Universal sobre los Archivos (ICA-UNESCO, 10 de noviembre de 2011)	Declaración del Consejo Internacional de Archivos adoptada por UNESCO	ica.org
Principio archivístico	Principios SAHR-ICA (2016)	Principios básicos sobre el papel de los archiveros y gestores de	ica.org

Categoría	Referente	Descripción	Acceso
		documentos en la defensa de los Derechos Humanos	
Referente nacional Colombia	**DANE** (2024) Guía para la anonimización de datos estructurados	Documento técnico de la Dirección de regulación, planeación, estandarización y normalización del DANE sobre anonimización de microdatos. Aplica al sistema estadístico nacional bajo Ley 2335 de 2023	dane.gov.co/files/sen/registros-administrativos/guia-anonimizacion-datos2024.pdf
Referente nacional Colombia	**CNMH** (2022) Guía para la anonimización de datos e información no estructurada	Complemento técnico de la guía del DANE: cubre datos no estructurados (texto libre, audio, video, imagen) en el contexto de archivos de DD. HH. y conflicto armado	centrodememoriahistorica.gov.co/wp-content/uploads/2022/08/GUIA-DE-ANONIMIZACION.pdf
Referente nacional Colombia	**AGN-CNMH** (2017, V.II 2022) Protocolo de gestión documental de los archivos referidos a las graves y manifiestas violaciones a los Derechos Humanos e infracciones al DIH ocurridas con ocasión del conflicto armado interno. Resolución conjunta 031 de 2017, actualizado en julio de 2022	De obligatoria adopción para entidades públicas, privadas con funciones públicas y demás organismos regulados por la Ley 594 de 2000	archivogeneral.gov.co/sites/default/files/Estructura_Web/5_Consulte/Recursos/Publicacionees/2022/protocolo-gd-ddhh.pdf
Referente nacional Colombia	**SIC (2021)** Guía sobre el tratamiento de datos personales en las entidades estatales	Marco institucional general bajo la Ley 1581 de 2012 y el Decreto 1074 de 2015 para el sector público	sic.gov.co/sites/default/files/files/2021/SIC%20(2021)%20Gu%C3%ADa%20sobre%20el%20tratamiento%20de%20datos%20personales%20en%20las%20entidades%20estatales.pdf

Fuente: elaboración propia

Marco legal de la anonimización en documentos de archivo

El marco legal de la anonimización en documentos de archivo se construye a partir de la confluencia de tres regímenes constitucionales y legales: el de transparencia y acceso a la información pública, el de protección de datos personales y el de archivos. La anonimización no es objeto principal de ninguno de los tres; es una herramienta cuya procedencia se determina a partir de su lectura armónica. El régimen de transparencia establece la regla de máxima publicidad y delimita los supuestos de clasificación y reserva; el de protección de datos define las categorías, principios de tratamiento y derechos de los titulares; y el de archivos garantiza la integridad, autenticidad y conservación del documento a lo largo de su ciclo de vida.

Sobre este marco general se superponen regímenes especiales que prevalecen en sus ámbitos específicos por razón de la materia, del sujeto protegido o del tipo de información tratada. Entre ellos se encuentran el régimen de inteligencia y contrainteligencia (Ley 1621 de 2013), el de protección reforzada de niños, niñas y adolescentes, en adelante, NNA, (Ley 1098 de 2006), el de justicia transicional y derechos de las víctimas (Ley 1448 de 2011 y normativa concordante de la JEP), el régimen estadístico (Ley 2335 de 2023) y el de reserva de la historia clínica (Ley 23 de 1981), entre otros. Algunos de estos regímenes introducen reglas específicas de clasificación, reserva, plazos y técnicas de protección que condicionan o desplazan la aplicación del marco general cuando el documento contiene información sometida a su ámbito de regulación. La anonimización no es objeto principal de ninguno de estos regímenes. Su procedencia, alcance y técnica admisible se determinan en cada caso concreto a partir de la calificación jurídica del documento, del tipo de información que contiene, del régimen aplicable por razón de la materia y del propósito al que se destina la divulgación o la consulta. Las secciones siguientes desarrollan, primero, los componentes del marco general y luego, los regímenes especiales.

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

El régimen general se construye a partir del artículo 15 de la Constitución Política, que reconoce el derecho fundamental al habeas data, y se desarrolla en la Ley 1581 de 2012, reglamentada por el Decreto 1074 de 2015. Esta ley contiene principios y disposiciones que sustentan decisiones técnicas en procesos de anonimización, sin que su sola existencia active un mandato general de anonimizar todo documento que contenga datos personales.

El artículo 4 consagra principios directamente conectados con la anonimización: el literal f) sobre circulación restringida exige que los datos personales no estén disponibles en internet u otros medios masivos salvo acceso técnicamente controlados; el literal g) sobre seguridad obliga a implementar medidas técnicas, humanas y administrativas para proteger los datos; y el literal h) sobre confidencialidad impone reserva permanente, incluso después de finalizada la relación con el tratamiento.

El artículo 6 prohíbe el tratamiento de datos sensibles, no obstante, el literal e) habilita el tratamiento con finalidad histórica, estadística o científica, condicionado a la supresión de la identidad de los titulares. Asimismo, el literal a) del artículo 10 habilita a las entidades públicas para tratar datos personales en ejercicio de sus funciones legales, sin autorización del titular ni anonimización previa.

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

El artículo 18 regula la información pública clasificada, exceptuada del acceso cuando su divulgación pueda afectar el derecho a la intimidad, incluida la de servidores públicos en los límites de su función, los derechos a la vida, salud o seguridad, y los secretos comerciales, industriales y profesionales. El parágrafo establece que estas excepciones tienen carácter indefinido, salvo consentimiento expreso del titular o cuando los datos fueron entregados bajo régimen de publicidad. La obligación de proteger estos datos persiste con independencia de la antigüedad del documento.

El artículo 19 regula la información pública reservada, cuyo acceso puede denegarse para proteger: la defensa y seguridad nacional, la seguridad pública, las relaciones internacionales, la prevención e investigación de delitos y faltas disciplinarias, el debido proceso e igualdad de las partes, la administración efectiva de justicia, los derechos de la infancia y adolescencia, la estabilidad macroeconómica y financiera, y la salud pública. A diferencia de la información clasificada, la reservada está sujeta a los plazos del artículo 22, transcurridos los cuales opera el levantamiento de la reserva. Sin embargo, el levantamiento no equivale a publicación íntegra: los datos personales privados y sensibles mantienen su protección.

El Decreto 103 de 2015, compilado en los Decretos 1081 y 1080 de 2015, establece que permitir el acceso a un dato semiprivado, privado o sensible no le quita su carácter clasificado ni implica desprotección; el acceso solo es posible por decisión de autoridad jurisdiccional, pública o

administrativa competente. Su artículo 31 introduce un criterio operativo central: cuando un documento contenga información mixta, divulgable y clasificada o reservada, el sujeto obligado debe revelar los datos no protegidos y justificar la reserva del resto. Para ello, autoriza expresamente el uso de tachado, anonimización, transliteración, edición y apertura de expedientes separados u otras medidas, según el formato y medio de conservación.

Ley 594 de 2000 - Ley General de Archivos

El artículo 4 consagra los principios de integridad, autenticidad, veracidad y conservación, conforme a los cuales los documentos deben preservarse en su forma original, sin alteraciones que comprometan su valor probatorio, testimonial o informativo. En materia de anonimización, esto implica que la intervención debe realizarse siempre sobre una copia de trabajo, conservando el original bajo condiciones de acceso restringido.

El artículo 27 establece el derecho de toda persona a consultar documentos de archivos públicos y obtener copia, salvo carácter reservado, e impone a las autoridades el deber simultáneo de garantizar la intimidad personal y familiar, la honra y el buen nombre.

Acuerdo 001 de 2024 del Archivo General de la Nación

El Acuerdo 001 de 2024 es el instrumento normativo más reciente en materia de gestión documental, de obligatorio cumplimiento para las entidades del Sistema Nacional de Archivos. En el contexto de la anonimización, establece las condiciones para la organización, descripción, acceso y transferencia documental, exigiendo que toda intervención sobre un documento quede debidamente registrada con trazabilidad institucional.

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 en su artículo 27 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.

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.

Tabla 3. Regímenes especiales, jurisprudencia constitucional y otros instrumentos internacionales

Categoría	Sector / Instrumento	Relación con la anonimización
Régimen especial de inteligencia y contrainteligencia	Ley 1621 de 2013, artículo 33: reserva legal de hasta treinta años y protección de la identidad de agentes, fuentes, medios y métodos. Decreto 1070 de 2015 y Decreto 857 de 2014: clasificación de documentos de inteligencia y protección de la identidad funcional de los agentes. Decreto 2149 de 2017: crea el Sistema Nacional de Depuración de Datos y Archivos de Inteligencia y Contrainteligencia	Imponen deberes de reserva y de protección de identidades cuya implementación requiere, en la práctica, técnicas de anonimización al levantar la reserva, depurar archivos o transferir documentos al patrimonio archivístico
Régimen especial para niños, niñas y adolescentes	Ley 1098 de 2006, artículo 33: derecho a la intimidad personal y a la protección frente a injerencias arbitrarias. Ley 1581 de 2012, artículo 7: tratamiento de datos personales de NNA bajo el principio de interés superior. Decreto 1377 de 2013, artículo 12 (compilado en Decreto 1074 de 2015): condiciones específicas de tratamiento	Establece el deber reforzado de proteger la intimidad y los datos personales de NNA, junto con el principio de interés superior, exige adoptar técnicas que impidan la identificación directa o indirecta cuando se accede, divulga o conserva información que les concierne
Régimen especial de Justicia transicional y derechos de las víctimas	Ley 1448 de 2011, Ley de Víctimas y Restitución de Tierras: protección de la información del Registro Único de Víctimas y de los archivos de derechos humanos. Jurisprudencia de la Jurisdicción Especial para la Paz (Autos 001, 073, 182 y 778): medidas cautelares sobre archivos del DAS y protocolos de acceso para el Sistema Integral de Verdad, Justicia, Reparación y No Repetición	Ordenan proteger la identidad de víctimas, testigos e informantes en escenarios donde el acceso a la información es esencial para garantizar el derecho a la verdad. La anonimización aparece como técnica idónea para conciliar ambos mandatos
Régimen especial de reserva estadística	Ley 2335 de 2023, "Por la cual se expiden disposiciones sobre las estadísticas oficiales en el país"	La ley regula la reserva estadística como deber del DANE y del

Categoría	Sector / Instrumento	Relación con la anonimización
		Sistema Estadístico Nacional, sin definir técnicas concretas. La anonimización de microdatos ha sido adoptada por el DANE como medio operativo para difundir información estadística sin vulnerar la reserva de la fuente
Régimen especial Reserva de la historia clínica	Ley 23 de 1981, artículo 33 y siguientes. Resolución 1995 de 1999 del Ministerio de Salud: manejo de la historia clínica. Ley 1581 de 2012, artículo 5: calificación de los datos de salud como datos sensibles	La normativa establece la reserva de la historia clínica y el carácter sensible de los datos de salud, pero no regula expresamente la anonimización. En la práctica, esta técnica permite habilitar el uso secundario de la información con fines de investigación, gestión sanitaria o estadística, respetando los deberes de reserva y confidencialidad.
Normativa archivística complementaria	Decreto 1400 de 2025: dispone que los documentos originales se mantengan inalterados durante el proceso de desclasificación del DAS. Decreto 1080 de 2015, Decreto Único Reglamentario del Sector Cultura, artículos 2.8.2.1.1 y siguientes: regulación del patrimonio archivístico	Sus disposiciones sobre integridad e inalterabilidad del documento original sustentan la regla técnica desarrollada en 5.3: cualquier intervención de anonimización debe realizarse sobre una copia de trabajo, preservando el original bajo régimen de acceso restringido
Jurisprudencia constitucional relevante	Sentencia T-729 de 2002: clasificación de los datos personales en públicos, semiprivados, privados y sensibles. Sentencia C-1011 de 2008: revisión de constitucionalidad de la Ley 1266 de 2008. Sentencia C-748 de 2011: revisión de constitucionalidad de la Ley 1581 de 2012; principio de interés superior en el tratamiento de datos de NNA. Sentencia C-274 de 2013: revisión de constitucionalidad de la Ley 1712 de 2014; regla de máxima publicidad y excepcionalidad de la reserva	La jurisprudencia constitucional no se pronuncia directamente sobre técnicas de anonimización, pero fija los estándares interpretativos sobre las categorías de datos personales y los supuestos de clasificación y reserva, criterios sustantivos a partir de los cuales se determina cuándo la anonimización resulta exigible y bajo qué condiciones
Estándares internacionales	Principios de Johannesburgo (1996) y Principios de Tshwane (2013): regla de máximo acceso con mínima restricción y protección de la identidad de víctimas, testigos e informantes. Conjunto actualizado	Ofrecen criterios interpretativos sobre el balance entre acceso y privacidad, la protección de identidades en archivos de derechos humanos y el principio de

Categoría	Sector / Instrumento	Relación con la anonimización
	de principios para la protección y promoción de los derechos humanos mediante la lucha contra la impunidad. Principios Joinet/Orentlicher (Doc. ONU E/CN.4/2005/102/Add.1, 2005). OECD Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data (1980, revisadas en 2013). Convenio 108+ del Consejo de Europa (modernizado en 2018). Resolución 68/167 de la Asamblea General de la ONU (2013), El derecho a la privacidad en la era digital	minimización del dato, que es el fundamento técnico-jurídico internacional sobre el que se sustenta la práctica de la anonimización

Fuente: elaboración propia.

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 reidentificación 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 reidentificación 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.

Nuevas formas de producción de información no estructurada

Las entidades deberán considerar que la información no estructurada puede originarse en entornos digitales contemporáneos, incluyendo, entre otros, sistemas de mensajería institucional, plataformas de videoconferencia, herramientas colaborativas, capturas de pantalla, así como contenidos generados mediante tecnologías de inteligencia artificial.

Estos formatos pueden contener datos personales o información sensible de manera implícita o contextual, por lo que su tratamiento requerirá la adopción de criterios técnicos y archivísticos adecuados para su identificación, gestión y eventual anonimización.

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. La inclusión de un formato en este listado no implica que su contenido deba anonimizarse; opera como referencia para identificar dónde pueden encontrarse este tipo de datos.

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 Word, PDF, informes, memorandos, oficios y actas; presentaciones institucionales (PowerPoint u otros formatos similares) que contengan texto libre, imágenes o datos personales integrados al contenido; 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, correos electrónicos incorporados formalmente al expediente como soporte probatorio o de gestión, y capturas de pantalla de aplicaciones, perfiles o interacciones digitales.

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 categorías de datos personales que pueden encontrarse dentro de los formatos no estructurados. Su presencia en un documento no implica por sí sola que deban anonimizarse; el inventario opera como referencia para su identificación, mientras que la decisión sobre la intervención se determina caso a caso, conforme al marco legal aplicable.

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. También comprenden registros de acceso físico a instalaciones, tarjetas de entrada y salida, y bitácoras de control de ingreso que aparecen frecuentemente en informes de seguridad institucional. Facilitan la identificación indirecta y el seguimiento de desplazamientos.

Identificadores documentales. Números de documento de identidad, NIT, tarjetas profesionales y registros civiles, a veces constituyen identificadores documentales más en archivos. 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. Metadatos incrustados en archivos digitales, incluyendo metadatos EXIF de fotografías (geolocalización, fecha, modelo del dispositivo, identidad del autor) e historial de revisiones y autoría en documentos de texto. Permiten la trazabilidad de actividades y la correlación de información entre sistemas.

Identificadores sensibles. El rostro y la voz constituyen los identificadores de mayor riesgo técnico en los formatos no estructurados, por cuanto permiten la identificación unívoca de las personas y son técnicamente difíciles de disociar del contenido del documento. A estos se suman otros identificadores presentes en documentos de archivo: huellas dactilares, iris, geometría de la mano y señales físicas visibles como tatuajes, cicatrices o marcas de nacimiento. Por su mayor dificultad técnica de disociación, las fotografías, videos, grabaciones de CCTV y registros de voz exigen, cuando proceda intervenirlos, técnicas especializadas aplicadas de manera consistente durante su

ciclo de uso y reproducción. La sola presencia de rostro o voz no activa por sí misma el deber de anonimizar: la decisión se determina caso a caso conforme a las categorías de sujetos protegidos, al régimen de publicidad aplicable y al análisis de riesgo contextual. Material audiovisual institucional como: videos de comunicaciones, audiencias públicas, fotografías de eventos oficiales o registros de actos administrativos abiertos al público, no requiere anonimización por la mera circunstancia de mostrar personas identificables.

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. A continuación, se presentan, a manera de ejemplo, las principales tipologías documentales. Su inclusión en este listado no implica que su contenido deba anonimizarse; opera como referencia para identificar dónde puede haber datos personales o información sensible. La decisión sobre la intervención se determina caso a caso, conforme al contexto, la finalidad del acceso y el marco normativo aplicable.

Documentos de gestión de personal y función pública. Expedientes de personal que incluyen hojas de vida, soportes de vinculación, evaluaciones, datos de contacto personal e 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 nacional 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, así como 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.

Para orientar el diagnóstico institucional sobre qué información requiere tratamiento, los presentes lineamientos establecen los criterios, las técnicas y los procedimientos que deben seguirse cuando la anonimización resulte procedente. Su aplicación inadecuada puede afectar la calidad de la información y el acceso efectivo a la misma, por lo que la regla general sigue siendo la máxima publicidad y la anonimización opera como excepción.

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.

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, conforme al artículo 3 literal h) de la Ley 1581 de 2012 y al artículo 3 del Decreto 1377 de 2013. Su tratamiento no requiere autorización del titular cuando se realiza en el marco de la finalidad institucional que justifica su carácter público. El derecho de oposición del titular sobre datos públicos opera de manera restringida: no procede frente a tratamientos amparados por habilitación legal o exigidos por la finalidad institucional que motivó su publicidad, pero sí cuando el tratamiento se desarrolle por fuera de esa finalidad o exceda los fines compatibles con la naturaleza pública del dato. 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 está prohibido por regla general, y requiere autorización expresa del titular. No pueden ser solicitados de manera obligatoria, salvo que una norma legal expresamente establezca esa obligación.

Finalmente, los datos de niños, niñas y adolescentes (NNA) merecen una mención especial por la protección constitucional reforzada que les otorga el ordenamiento jurídico, en aplicación del artículo 44 de la Constitución, el artículo 7 de la Ley 1581 de 2012 y la Sentencia C-748 de 2011 de la Corte Constitucional. El tratamiento debe atender en todo caso al interés superior del menor y al respeto de sus derechos fundamentales. La regla de autorización opera así: cuando el tratamiento se realiza en ejercicio de funciones legales de la entidad pública, por orden judicial o en virtud de habilitación legal expresa, supuestos del artículo 10 de la Ley 1581 de 2012, no se requiere autorización del representante legal, sin perjuicio del deber de adoptar medidas reforzadas de protección y de motivar la decisión conforme al interés superior del menor. Cuando no exista habilitación legal, el tratamiento de datos públicos y semiprivados de NNA requiere autorización del representante legal; el tratamiento de datos privados y sensibles de NNA no debe realizarse, salvo que ello resulte indispensable para garantizar los derechos del menor, hipótesis en la cual deberá contarse con autorización cualificada del representante legal y, cuando proceda, con la opinión del menor según su grado de madurez. 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.

Identificadores directos e indirectos

Conforme al desarrollo conceptual contenido en la Guía de Anonimización de Datos Estructurados del AGN (2020) y en la Guía para la anonimización de datos estructurados del DANE (2024), 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.

La Ley 1581 de 2012 define el dato personal como "cualquier información vinculada o que pueda asociarse a una o varias personas naturales determinadas o determinables" (artículo 3, literal c). El término determinable es el criterio jurídico decisivo: un dato es personal no solo cuando permite identificar o determinar a alguien de manera inmediata, sino cuando esa identificación o determinación es posible, incluso si requiere pasos adicionales o el cruce con otras fuentes. En ese sentido, por ejemplo, la clasificación legal de un dato como público no equivale a su libre tratamiento ni a la ausencia de riesgo.

Por estas razones, la tabla que se presenta a continuación debe leerse como una guía orientadora y no como un criterio concluyente. La clasificación jurídica de cada dato es el punto de partida del análisis, no su resultado. La decisión sobre el alcance de la anonimización requerida en cada caso debe incorporar, además de esa clasificación, una valoración del contexto documental, la finalidad del acceso, el riesgo de reidentificación dado el estado actual de la tecnología, y la posibilidad de que el documento permita inferir atributos sensibles, aunque no los contenga de manera explícita. Esa valoración no puede ser estática: debe revisarse periódicamente conforme evolucionen las capacidades tecnológicas disponibles para el tratamiento de la información.

Tabla 4. Categorías de datos personales según su nivel de protección[10]

Tipo de datos personales	Variación contextual del nivel de protección
Nombres y apellidos	Adquiere carácter privado o sensible cuando aparece en expedientes disciplinarios, registros de víctimas, actuaciones judiciales, informes de inteligencia o cualquier documento que revele una situación de vulnerabilidad, riesgo o estigmatización

10 Nota sobre la lectura de esta tabla. La clasificación de cada dato como público, semiprivado, privado o sensible no es fija. La columna "Variación contextual del nivel de protección" identifica los principales escenarios en los que un dato escala hacia mayor protección por efecto del contexto documental. La protección también puede atenuarse cuando el dato pierde su capacidad identificadora, por ejemplo, una vez transcurrido el plazo de reserva, cuando ha sido publicado lícitamente por el titular en ejercicio de su autodeterminación informativa, o cuando el documento ha sido sometido a anonimización efectiva conforme al numeral 9 de este lineamiento. En ambos sentidos, la evaluación corresponde al responsable del tratamiento y debe quedar documentada en el expediente, sin que la inclusión genérica de un dato en una categoría exima del análisis caso por caso.

Tipo de datos personales	Variación contextual del nivel de protección
Tipo de identificación, número de identificación	Adquiere carácter privado cuando se combina con datos de salud, situación judicial, condición de víctima o cualquier atributo que permita perfilar a la persona más allá de su identidad formal
Fecha y lugar de expedición	Puede combinarse con otros cuasi-identificadores del mismo documento para permitir la individualización de la persona, especialmente en poblaciones pequeñas o comunidades étnicas
Estado civil	Adquiere carácter sensible cuando aparece en documentos que involucran procesos de adopción, violencia intrafamiliar, reconocimiento de uniones de hecho de personas del mismo sexo o situaciones de vulnerabilidad familiar
Sexo	Adquiere carácter sensible cuando su combinación con otros datos del documento permite inferir identidad de género u orientación sexual, especialmente en contextos de violencia, discriminación o actuaciones judiciales
Nacionalidad	Adquiere carácter sensible cuando se combina con datos étnicos, de desplazamiento o de condición migratoria irregular, o cuando permite identificar a una persona dentro de una comunidad minoritaria específica
Correo electrónico institucional-corporativo	Por regla general es información pública en virtud de la Ley 1712 de 2014. Solo adquiere carácter privado de manera excepcional cuando su divulgación expone al funcionario a un riesgo cierto para su vida, salud o seguridad debidamente acreditado, o cuando el cargo asociado se desempeña en contextos de inteligencia, contrainteligencia, seguridad del Estado. La identificación de la estructura organizacional interna no es causal válida de anonimización: el organigrama y los canales institucionales de contacto son objeto de publicación proactiva conforme a la Ley 1712 de 2014
Teléfono y celular corporativo	Por regla general es información pública conforme a la Ley 1712 de 2014. Solo adquiere carácter privado de manera excepcional cuando su publicación expone al funcionario a un riesgo cierto para su vida, salud o seguridad debidamente acreditado, o cuando el cargo asociado implica funciones de seguridad, protección, inteligencia o contrainteligencia. La mera posibilidad de contacto no deseado o de incomodidad institucional no constituye una causal de anonimización
Dirección de trabajo	Adquiere carácter privado o sensible en documentos de inteligencia, expedientes de víctimas o actuaciones que involucren personas en riesgo, donde la ubicación laboral puede facilitar su localización
Datos relativos a la actividad comercial	Adquiere carácter semiprivado o privado cuando su combinación con otros datos del expediente permite inferir la situación económica, patrimonial o de vulnerabilidad de una persona natural
Datos relativos a la actividad profesional	Adquiere carácter privado cuando aparece en actuaciones disciplinarias, procesos de responsabilidad fiscal o documentos que asocien la actividad profesional con conductas sancionables o situaciones de riesgo

Tipo de datos personales	Variación contextual del nivel de protección
Datos generales relacionados con afiliación al sistema integral de seguridad social en salud (EPS)	Adquiere carácter sensible cuando su combinación con otros datos del documento permite inferir una condición de salud, una incapacidad o una situación de discapacidad de la persona
Lugar y fecha de nacimiento	Constituye un cuasi-identificador de alto riesgo en poblaciones pequeñas o comunidades étnicas, donde su combinación con otros atributos puede permitir la identificación unívoca de la persona
Edad	Adquiere carácter sensible cuando permite identificar a una persona como menor de edad en actuaciones judiciales o administrativas, activando la protección reforzada de la Ley 1098 de 2006
Correo electrónico personal	Adquiere carácter privado cuando su divulgación permite inferir vínculos personales, afiliaciones políticas o religiosas, o cuando aparece en documentos que involucran situaciones de vulnerabilidad o riesgo
Celular y teléfono personal	Adquiere carácter privado o sensible en expedientes de víctimas, actuaciones de inteligencia o documentos en los que la localización de la persona represente un riesgo para su seguridad
Datos financieros, crediticios y/o derechos de carácter económico y datos patrimoniales	Adquiere carácter privado cuando su divulgación afecta la dignidad económica de la persona o permite inferir condiciones de vulnerabilidad, endeudamiento o precariedad
Datos socioeconómicos: estrato, propiedad de vivienda, bienes muebles e inmuebles, ingresos, egresos, inversiones	Adquiere carácter sensible cuando su combinación con datos étnicos, de desplazamiento o de victimización permite perfilar a una comunidad o grupo en situación de vulnerabilidad estructural
Datos de información tributaria	Adquiere carácter privado cuando su divulgación excede la finalidad de control fiscal y permite inferir la situación patrimonial o económica de una persona natural en contextos ajenos al control público
Datos relacionados con la actividad económica de la persona	Adquiere carácter privado o sensible cuando permite inferir vínculos con actividades ilícitas, grupos armados o redes de financiamiento en contextos de seguridad o justicia transicional
Datos relacionados con la historia laboral: experiencia, cargo, fechas de ingreso y retiro, anotaciones, llamados de atención	Adquiere carácter privado cuando aparece en actuaciones disciplinarias o sancionatorias, y carácter sensible cuando permite inferir afiliación sindical, activismo político o condición de víctima en el lugar de trabajo
Datos relacionados con el nivel educativo, capacitación e historial académico	Adquiere carácter sensible cuando su combinación con datos étnicos o de discapacidad permite perfilar condiciones de discriminación o exclusión estructural
Datos de antecedentes judiciales y/o disciplinarios	Adquiere carácter sensible cuando involucra a personas que han sido víctimas del sistema penal o disciplinario en contextos de persecución

Tipo de datos personales	Variación contextual del nivel de protección
	política, sindical o étnica, o cuando su divulgación puede generar estigmatización
Datos personales de acceso a sistemas de información: usuarios, IP, claves, perfiles	Adquiere carácter sensible cuando permite inferir patrones de comportamiento, afiliaciones ideológicas o actividades de la persona en entornos digitales institucionales o personales
Imagen (fotografía y video)	Su clasificación depende del contenido y el contexto: una fotografía institucional puede ser pública; una imagen que revele condiciones de salud, origen étnico, situación de víctima o contexto de riesgo tiene carácter sensible. Cuando permite identificación automatizada debe tratarse como dato biométrico. La evaluación debe hacerse documento por documento
Datos sensibles: huella, ADN, voz, cuando permitan identificar a una persona ante un sistema de información	Excepción a la dinamicidad contextual. Por su carácter de información que identifica de manera unívoca y permanente al titular, conservan condición sensible en todo contexto documental sin admitir graduación. Su tratamiento requiere autorización previa, expresa e informada del titular y no se beneficia de los supuestos de atenuación aplicables a otras categorías. La anonimización efectiva exige disociación irreversible o supresión, no siendo suficiente la sustitución o el control de acceso
Datos de descripción morfológica: color de piel, iris, cabello, señales particulares, estatura, peso, complexión	Adquiere carácter privado o sensible cuando, por su nivel de detalle, permite individualizar al sujeto o asociarlo a una condición de salud, discapacidad, origen étnico u otra categoría especialmente protegida. En documentos narrativos, judiciales o administrativos puede funcionar como identificador indirecto, especialmente cuando se combina con referencias territoriales o temporales específicas. Su protección se atenúa cuando la descripción es genérica y no permite individualizar al sujeto en el universo poblacional pertinente
Datos relacionados con la salud de la persona	Pueden aparecer de forma implícita en narrativas que describen situaciones de incapacidad, ausencia laboral o limitaciones funcionales sin mencionar diagnósticos explícitos. El dato inferido tiene el mismo nivel de protección que el dato explícito
Datos sobre el estado de salud que incluyan resultados de pruebas, laboratorios, diagnósticos médicos, psicológicos, psiquiátricos, medicamentos y tratamientos terapéuticos	En expedientes de personal, actuaciones disciplinarias o registros de víctimas, estos datos pueden aparecer dispersos en el texto narrativo sin estar formalmente identificados como información médica. Requieren revisión humana y contextual
Datos relacionados con pertenencia a sindicatos, organizaciones sociales, de derechos humanos, religiosas o políticas; convicciones religiosas, filosóficas o políticas	En archivos del conflicto armado, informes de inteligencia o expedientes disciplinarios, estos datos pueden aparecer de forma implícita en descripciones de actividades, reuniones o vínculos asociativos. Su inferencia a partir del texto tiene el mismo nivel de protección

Tipo de datos personales	Variación contextual del nivel de protección
Datos de preferencia, identidad y orientación sexual; origen étnico-racial	Son especialmente vulnerables a la inferencia en documentos narrativos. Una descripción de hechos, un relato de víctima o un informe de campo puede revelar estos atributos sin nombrarlos explícitamente. Requieren análisis contextual riguroso
Población en condición de vulnerabilidad	La condición de víctima puede inferirse de la narrativa documental, aunque el documento no lo declare formalmente. En archivos del conflicto armado, cualquier dato asociado a una persona identificable debe tratarse bajo este estándar de protección
Datos sobre personas en situación de discapacidad	Pueden aparecer de forma implícita en descripciones de limitaciones funcionales, referencias a apoyos o ajustes razonables, o menciones de condiciones físicas o cognitivas en expedientes administrativos o judiciales

Fuente: elaboración propia.

La tabla anterior debe leerse de manera articulada con la tabla de identificadores directos e indirectos contenida en la Guía de Anonimización de Datos Estructurados del AGN (2020). Mientras la tabla de esta sección clasifica los datos personales según su nivel de protección jurídica, e incorpora advertencias sobre su posible variación, 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.

En el caso específico de información no estructurada en formato audiovisual y multimedia, los identificadores presentes pueden clasificarse en tres categorías diferenciadas, conforme al desarrollo conceptual de la Guía para la Anonimización de Datos e Información No Estructurada (CNMH-AGN, 2022, numeral 10.3): i) identificadores no biométricos, que incluyen el contexto del texto, el contexto del discurso, las matrículas de vehículos e inmuebles, el contexto sociopolítico y ambiental específico, el estilo de vestir y el peinado; ii) identificadores biométricos, esto es, las características personales distintivas, mensurables, generalmente únicas y permanentes, clasificadas como fisiológicas (rostro, iris, oído, huella digital) y conductuales (voz, marcha, gesto, movimiento de labios, estilo de mecanografía), y iii) identificadores biométricos blandos, que comprenden características físicas, de comportamiento o adheridas, vagas y no necesariamente permanentes (altura, peso, color de ojos, silueta, edad, sexo, lunares, tatuajes, marcas de nacimiento, cicatrices). Aunque los identificadores biométricos blandos por sí solos no permiten una identificación personal confiable, pueden combinarse con otros identificadores para reforzar el reconocimiento o para clasificar a las personas en categorías particulares, lo que también puede resultar invasivo para la privacidad. La intervención técnica sobre material audiovisual debe

considerar las tres categorías y no limitarse a la supresión de los identificadores fisiológicos evidentes.

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

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. A continuación, se describen los principales sujetos que deben considerarse en el proceso de anonimización:

Servidores públicos y exservidores públicos. Por regla general, los datos personales del servidor público asociados al ejercicio de sus funciones y demás información ligada a su gestión pública, son información pública y no deben anonimizarse, en aplicación del principio de máxima publicidad. El carácter público del cargo conlleva una expectativa reducida de intimidad, en concordancia con el

artículo 18 de la Ley 1712 de 2014. La anonimización procede únicamente de manera excepcional cuando sus datos personales estén asociados de su esfera estrictamente personal o familiar y no estén vinculados al ejercicio del cargo; o cuando la identificación del servidor exponga indirectamente a terceros protegidos; o exista cierto riesgo para la vida, la salud o la seguridad del servidor o de su familia y se encuentre debidamente acreditado. No se justifica por sí sola la anonimización del cargo, la condición de exservidor, el deseo de reserva del titular o la incomodidad institucional frente al control ciudadano.

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 NNA en cualquier actuación, a los niños, niñas y adolescentes intersexuales, y a sus representantes legales, defensores y agentes oficiosos cuando su identificación permita inferir la identidad del menor. También deben anonimizarse las juezas, jueces, personal médico tratante u otros funcionarios cuando su identificación exponga indirectamente al menor involucrado.

Personas en situación 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 personas migrantes, refugiadas, desplazadas, comunidades indígenas y familias pertenecientes a grupos étnicos.

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.

Denunciantes y alertadores. Personas que reportan conductas irregulares, actos de corrupción o violaciones a derechos humanos en el marco de actuaciones de control interno, disciplinario, fiscal o penal. Su identificación en expedientes administrativos o judiciales puede generar riesgos de represalia, intimidación o afectación laboral.

Funcionarios internacionales y personal diplomático. Aparecen en archivos de la Cancillería, acuerdos internacionales y documentos de cooperación. Su identificación puede tener implicaciones en las relaciones internacionales.

Espacios, lugares y hechos objeto de protección

Con el fin de complementar la identificación de sujetos de protección, esta sección establece los lugares, espacios físicos y hechos 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. 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 niños, niñas y adolescentes (NNA); las comisarías de familia en procesos relacionados con NNA o violencia intrafamiliar; los juzgados de familia cuando el proceso involucra niños, niñas y adolescentes; las estaciones de policía en casos que involucren NNA o víctimas de violencia; los juzgados penales cuando la identificación del despacho permita inferir la identidad de NNA o víctimas; las autoridades judiciales de instancia cuando el asunto involucre NNA 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 de registros de expedientes administrativos. 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. Esta protección opera, como regla general, mientras la investigación se encuentre en curso o la decisión no se halle en firme, en aplicación de la presunción de inocencia y de la protección del buen nombre y la intimidad del investigado.

Espacios vinculados a hechos 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 situación de vulnerabilidad de los titulares y deben ser tratados con el más alto nivel de protección.

Hechos 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 o incidentes internos que pueden afectar la honra y el buen nombre de los involucrados antes de que exista una decisión en firme.

En cuanto a los eventos históricos de alto impacto, cuando el hecho, la fecha y el lugar son de conocimiento público (como ocurre, por ejemplo, con la Toma del Palacio de Justicia), dicha información debe conservarse. Respecto de las personas, el criterio aplicable no es el de "figura pública", insuficiente y a veces contraproducente en contextos de memoria, sino el conjunto de criterios desarrollados en el lineamiento sobre espacios de memoria histórica: procede mantener la identificación de las víctimas cuando han consentido la divulgación, cuando ya han sido identificadas en escenarios oficiales (sentencias, decisiones de la JEP, informes de la Comisión del Esclarecimiento de la Verdad, decisiones de la Corte Interamericana de Derechos Humanos) o cuando su identificación resulta necesaria para la comprensión del hecho y para el ejercicio del derecho colectivo a la memoria. Procede la anonimización, en cambio, sobre terceros incidentales no consentidos, datos sensibles de la víctima ajenos al hecho memorable, presuntos victimarios sin condena en firme y personas vivas frente a las que existan riesgos actuales y acreditados.

La determinación de qué personas resultan necesarias para la comprensión histórica de un evento, cuáles han sido ya identificadas en escenarios oficiales y cuáles deben ser objeto de anonimización, no puede ser una decisión unilateral del jefe de oficina ni de la persona responsable individual del trámite. La complejidad de esta valoración, que articula derechos de las víctimas, derecho colectivo a la memoria, intimidad de terceros y eventuales riesgos actuales, exige un

análisis colegiado e interdisciplinar con competencias archivísticas, históricas y jurídicas, idealmente con participación o consulta de las víctimas o sus organizaciones representativas cuando ello sea posible.

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.

Técnicas de anonimización para información no estructurada

La anonimización en información no estructurada no se limita a una única técnica; la entidad puede aplicar distintas modalidades de intervención según el soporte, el nivel de riesgo, la finalidad de divulgación y el marco normativo aplicable a la tipología documental. La descripción de las técnicas que se presenta a continuación no implica que todo documento que contenga datos personales deba intervenirse. La procedencia de la anonimización se determina caso a caso. La selección de la técnica adecuada, una vez determinada la procedencia, 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[11].

11 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

El catálogo de técnicas que se desarrolla en los numerales siguientes está diseñado para la intervención sobre información no estructurada. Cuando el documento de archivo objeto de intervención contenga componentes estructurados, como por ejemplo tablas embebidas con datos personales, bases de datos anexas o registros tabulares incorporados al expediente, las técnicas aplicables a esos componentes son las descritas en la Guía para la anonimización de datos estructurados del DANE (2024), entre ellas microagregación, recodificación global, supresión local, intercambio de datos, codificación superior e inferior, redondeo y adición de ruido. La intervención sobre componentes estructurados y no estructurados de un mismo expediente debe realizarse de forma articulada, asegurando la coherencia de la versión pública resultante.

Regla universal: copia de trabajo. Principio de integridad del original. Toda intervención de anonimización debe realizarse sobre una copia de trabajo del documento, manteniendo el original íntegro, inalterado y bajo condiciones de custodia y acceso restringido. Este principio es de aplicación universal en el cumplimiento de los principios de integridad, autenticidad y conservación. Toda intervención de anonimización debe realizarse sobre una copia de trabajo del documento, manteniendo el original íntegro, inalterado y bajo condiciones de custodia y acceso restringido. Esta regla es de aplicación universal en el cumplimiento de los principios de integridad, autenticidad y conservación consagrados en el artículo 4 de la Ley 594 de 2000 y en el Acuerdo 001 de 2024 del AGN, y se aplica con especial rigor cuando se trate de archivos referidos a derechos humanos y derecho internacional humanitario, archivos de memoria histórica o cualquier acervo documental con valor probatorio o testimonial.

Clasificación funcional de las técnicas

Las técnicas que se describen a continuación pueden agruparse, según su lógica de intervención sobre el dato, en cinco familias funcionales reconocidas por los estándares internacionales de referencia[12]:

i. **Reducción de información.** La intervención consiste en eliminar o disminuir el contenido informativo del documento sin alterar lo que permanece. Comprende la supresión o eliminación directa, el enmascaramiento de caracteres y, en su modalidad de reducción del nivel de detalle, la generalización.
ii. **Sustitución y disociación.** La intervención consiste en reemplazar el dato identificador por un valor neutro o en romper el vínculo entre el dato y su titular. Comprende la sustitución por etiquetas, la codificación y tokenización semántica, la disociación irreversible y la separación de soportes o bifurcación del expediente. En cuanto a la disociación, esta

12 ISO/IEC 20889:2018; Article 29 Working Party, Dictamen 05/2014; Agencia Española de Protección de Datos, 2016.

intervención consiste en romper de manera irreversible el vínculo entre el dato y su titular. Comprende la disociación irreversible mediante destrucción de claves o tablas de correspondencia. La sustitución por etiquetas y la codificación o tokenización semántica con glosario interno, aunque reemplazan identificadores, conservan reversibilidad y se ubican por ello en la familia v (Control de acceso). La separación de soportes o bifurcación del expediente no es una técnica autónoma, sino un componente operativo de combinaciones

iii. **Aleatorización y perturbación.** La intervención consiste en introducir variaciones controladas que conservan utilidad estadística, pero rompen la correspondencia exacta con el dato original.

iv. **Reformulación del contenido.** La intervención consiste en producir una versión derivada del documento que conserva el sentido, pero no los identificadores. Comprende el resumen o abstracción y la agregación estadística.

Esta clasificación es funcional y no es excluyente. En la práctica, una intervención de anonimización suele combinar técnicas de varias familias. La selección depende del soporte, el nivel de riesgo, la finalidad de divulgación y el contexto archivístico, no de una preferencia abstracta por una familia sobre otra.

Supresión o eliminación directa

La supresión o eliminación directa consiste en retirar completamente del contenido el dato identificador, sin sustituirlo por otra referencia[13]. 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 crítica en esta técnica 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 soporte físico, debe emplearse tinta

13 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

opaca permanente aplicada sobre el dato verificando que no sea legible bajo luz ultravioleta o mediante escaneo de alta resolución.

Para soportes físicos el procedimiento recomendado de supresión mediante tachado sigue el proceso de doble copia: en primer lugar, se obtiene una copia del documento original mediante reprografía; sobre esa primera copia se realiza el tachado con marcador permanente de tinta opaca; de la primera copia se obtiene una segunda copia, que constituye la versión anonimizada del documento y es la única que puede divulgarse o entregarse. El documento original se conserva sin ninguna intervención, tal como fue recibido o producido. La primera copia, intervenida con el tachado, debe eliminarse de manera segura una vez generada la segunda copia, siguiendo los protocolos de destrucción documental que correspondan.

Cuando la información a suprimir corresponde a párrafos completos en soporte físico, puede cubrirse con una hoja en blanco antes de la reproducción, generando así la segunda copia sin necesidad de intervenir directamente el papel del documento original. Desde el punto de vista operativo, estos procesos no deben realizarse en lugares de acceso público ni en equipos fuera de la entidad, con el fin de evitar que la información pueda ser restaurada o visualizada por terceros durante el proceso de supresión.

La versión anonimizada resultante (segunda copia) debe conservarse con los códigos y la signatura topográfica original del expediente (fondo, serie, subserie, caja, carpeta) en un expediente nuevo o derivado, para asegurar la trazabilidad del procedimiento y garantizar el acceso si la información es requerida nuevamente[14].

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*

14 Centro Nacional de Memoria Histórica / Archivo General de la Nación. Guía para la Anonimización de Datos e Información No Estructurada: Estándares y Lineamientos Técnicos. Bogotá: CNMH, 2022. Numerales 9.1, 10.2, 10.3, 10.4 y Anexo 1.

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.

Sustitución por etiquetas, codificación y tokenización semántica

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.

Esta sección agrupa dos modalidades operativas que comparten la lógica de reemplazar identificadores por códigos manteniendo bajo reserva la correspondencia con la identidad real: la sustitución por etiquetas con glosario interno (descrita a continuación) y la seudonimización controlada con tabla de equivalencias. Ambas modalidades preservan la reversibilidad del proceso por parte de la entidad responsable y, por esa razón, no constituyen anonimización en sentido estricto; el punto 9.2.2 desarrolla los criterios técnicos que permiten distinguir cuándo se trata de una intervención efectivamente anonimiza y cuándo opera únicamente como mecanismo de gestión interna.

Un riesgo principal de esta técnica en información no estructurada es la inconsistencia entre documentos de un mismo expediente: si un mismo sujeto recibe códigos distintos en diferentes piezas procesales, la correlación entre documentos puede permitir su reidentificación. Por esta razón, el sistema de codificación debe establecerse desde el inicio del proceso y aplicarse de manera uniforme en todas las piezas documentales, incluidos los anexos, las pruebas y las comunicaciones asociadas. 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.

Esta técnica puede implementarse en dos modalidades según el volumen documental y la capacidad técnica de la entidad. La primera es la modalidad manual, adecuada para expedientes individuales o corpus de tamaño reducido, en la que el equipo responsable identifica y sustituye los identificadores de manera directa sobre el documento. La segunda es la modalidad automatizada o tokenización semántica, aplicable a corpus documentales extensos, en la que se utilizan herramientas de reconocimiento de entidades nombradas (NER, por sus siglas en inglés) que identifican y reemplazan automáticamente personas, lugares, organizaciones, fechas y números de identificación por tokens genéricos.

Independientemente de la modalidad utilizada, la precisión del resultado debe verificarse mediante revisión humana antes de la divulgación. En la modalidad automatizada esto es especialmente importante: en el caso del español colombiano, con sus particularidades léxicas, topónimos locales y nombres propios regionales, los modelos genéricos pueden presentar tasas de error significativas, dejando entidades sin identificar o tokenizando erróneamente elementos que no corresponden a datos personales. La revisión humana no es opcional sino parte integral del proceso.

Ejemplo de aplicación: *Es común en providencias judiciales que involucran 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.

Seudonimización controlada

La seudonimización 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 equivalencias[15]. Su aplicación es adecuada en investigaciones, análisis histórico o proyectos de interoperabilidad, y 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 la seudonimización en una medida insuficiente.

Desde el punto de vista operativo, los seudónimos deben diseñarse con la misma longitud y formato para facilitar la legibilidad y el procesamiento sistemático. Adicionalmente, los seudónimos generados para uso externo deben ser diferentes a los generados para uso interno, sin relación morfológica entre unos y otros, de modo que el conocimiento de un seudónimo externo no permita inferir el seudónimo interno ni la identidad real del titular.

Debe tenerse en cuenta que la seudonimización 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, la seudonimización debe combinarse con generalización de los atributos contextuales cuando el riesgo de correlación sea significativo.

15 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, 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.*

Criterios para distinguir la anonimización de la seudonimización

La confusión entre ambas figuras es frecuente y puede generar la falsa percepción de un nivel de protección que no se cumple. Para verificar si una intervención técnica alcanzó la anonimización o se quedó en la seudonimización, la entidad debe aplicar tres criterios concurrentes:

- **Reversibilidad**. Si el proceso puede revertirse mediante clave, tabla de equivalencias, llave criptográfica o cualquier mecanismo que permita reconstruir la identidad del titular, se trata de seudonimización. La anonimización exige que la reversibilidad sea técnicamente imposible o no factible con medios razonablemente disponibles para terceros.
- **Singularización, vinculabilidad e inferencia.** Conforme al Dictamen 05/2014 del Article 29 Working Party, la anonimización exige neutralizar simultáneamente los tres riesgos: aislar a un individuo en el conjunto, vincular registros referidos a la misma persona y deducir atributos a partir de los demás. Si cualquiera persiste, la intervención es seudonimización.
- **Universo poblacional.** En contextos archivísticos con universos reducidos (entidades pequeñas, series específicas, localidades poco pobladas), el umbral es más exigente: la sustitución de identificadores directos puede ser insuficiente cuando la combinación de atributos contextuales restantes (cargo, fecha, dependencia, hechos) reduce el universo a un grupo identificable.

Los tres criterios son acumulativos; una deficiencia en cualquiera sitúa la intervención en el plano de la seudonimización.

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 titular[16]. 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)
- **Código postal:** código postal completo → municipio → región
- **Ocupación/sector laboral:** cargo específico → sector → categoría general (empleado, independiente, desempleado)
- **Nivel de escolaridad:** título exacto → nivel educativo (básica, media, técnica, universitaria, posgrado).

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 una persona si el universo de personas con esas características es reducido. En estos casos, debe aplicarse un nivel de generalización adicional o combinarse con supresión.

16 Rocher, L., Hendrickx, J., y 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

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.*

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 advertencia técnica relevante en información no estructurada es el riesgo de referencias cruzadas internas: en un texto narrativo puede suprimirse el nombre de una persona en un párrafo, pero si en otro párrafo del mismo documento se hace referencia a "el funcionario mencionado anteriormente" o "la persona descrita en el apartado precedente", la disociación es incompleta. Antes de dar por concluido el proceso de disociación, debe realizarse una revisión integral del documento para identificar y tratar todas las referencias directas e indirectas al titular, incluyendo pronombres, cargos, apodos y cualquier otra referencia que permita vincular los párrafos entre sí.

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.*

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 voz[17]. 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.

Un riesgo frecuentemente subestimado en grabaciones de interiores institucionales es la reidentificación por fondo visual: elementos del entorno como cuadros institucionales, diplomas,

17 Todt, J., Hanisch, S., & Strufe, T. (2024). Fantômas: Understanding Face Anonymization Reversibility. Disponible en: https://www.researchgate.net/publication/239665775

carteles, mobiliario específico o avisos pueden identificar el espacio físico y, por tanto, a las personas que allí trabajan o son atendidas. Antes de divulgar grabaciones de audiencias, entrevistas o actuaciones institucionales, debe revisarse que el fondo no contenga elementos identificadores del espacio o de las personas presentes.

En cuanto 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.

Esta técnica aplica también a documentos escaneados que contengan imágenes de rostros incrustadas, como hojas de vida digitalizadas con fotografía, carnés escaneados o fichas técnicas con imagen. En estos casos, la distorsión técnica debe aplicarse sobre la imagen incrustada antes de la divulgación del documento, con independencia de que el resto del contenido haya sido tratado mediante otras técnicas.

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.

Adicionalmente a las técnicas de distorsión aplicadas en posproducción, los lineamientos contemplan, como medidas equivalentes en su efecto desidentificador, las técnicas de reinterpretación, dramatización, relectura y reconstrucción audiovisual. Estas consisten en producir una versión derivada del material original en la que los testimonios, hechos o relatos son representados por actores, narradores o intérpretes distintos a los sujetos originales, o son recreados en locaciones neutras y con escenografía no reconocible, conservando el contenido sustantivo, pero rompiendo la correspondencia entre la representación pública y la identidad de los titulares. Estas técnicas resultan especialmente útiles cuando la distorsión técnica de la imagen o la voz comprometería el valor narrativo, testimonial o pedagógico del material, o cuando el material ha de ser utilizado en escenarios de divulgación pública amplia (exposiciones, espacios de memoria, productos audiovisuales de divulgación). Su aplicación requiere autorizaciones de uso, sesiones de derechos sobre las locaciones, y la documentación expresa de la naturaleza derivada del material, de modo que el público destinatario comprenda que se trata de una representación y no del registro original. La experiencia institucional acumulada por el Centro Nacional de Memoria Histórica en

proyectos de divulgación con base en archivos del conflicto armado constituye una referencia operativa relevante para el diseño y aplicación de estas técnicas.

> ***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, que el fondo visual no contenga elementos que identifiquen el espacio o a las personas presentes, y que los metadatos del archivo hayan sido eliminados antes de su publicación o entrega.*

Enmascaramiento de caracteres

El enmascaramiento de caracteres consiste en ocultar parcialmente el valor de un atributo reemplazando algunos de sus caracteres por un símbolo neutro, por ejemplo, un asterisco o una equis, conservando los demás. El número de cédula 1234567631 se presenta como 1234****31. A diferencia de la supresión total, el enmascaramiento permite confirmar la existencia del dato sin revelar su valor completo.

Esta técnica es útil cuando la finalidad del documento exige acreditar que el dato existe, pero no es necesario conocer su contenido íntegro. Su aplicación debe limitarse a identificadores directos como números de cédula, números de teléfono, cuentas bancarias, códigos de afiliación al sistema de salud y seguridad social, y no es suficiente por sí sola en contextos de alta sensibilidad o publicación masiva, en los que debe combinarse con otras técnicas como la supresión o la generalización[18].

> ***Ejemplo de aplicación:*** *En la publicación de actos administrativos que contienen números de identificación de personas naturales, puede enmascararse la parte central del número conservando los primeros y los últimos dígitos, de modo que sea posible verificar la existencia del dato sin exponer el valor completo. Esta técnica no es adecuada como medida única en documentos de acceso masivo o*

18 Unidad para la Atención y Reparación Integral a las Víctimas. Lineamiento de Anonimización de Datos. Bogotá. UARIV, Oficina de Tecnologías de la Información, 2025.

en contextos en que el número de documento pueda ser inferido por correlación con otros atributos del titular.

Resumen o abstracción

El resumen o abstracción consiste en reemplazar el documento original o secciones de él por una versión resumida o abstracta que capture la información pública relevante sin reproducir los datos personales contenidos en el original. A diferencia del acceso parcial o mediado, que restringe lo que se entrega del documento existente, el resumen genera un documento nuevo que describe el contenido sin reproducirlo.

Esta técnica es aplicable cuando el documento contiene información de alto valor histórico, informativo o probatorio, pero también datos de alta sensibilidad que no pueden suprimirse sin afectar la coherencia del texto original. En estos casos, publicar un resumen descriptivo del expediente (indicando su objeto, las partes involucradas de manera genérica, el resultado de la actuación y su relevancia pública) permite cumplir con las obligaciones de transparencia sin exponer los datos personales que el documento contiene.

En la práctica archivística, esta técnica es especialmente útil para la publicación de instrumentos de descripción documental (inventarios, catálogos, guías de fondo) que describen el contenido de los expedientes sin reproducirlo. La descripción archivística puede diseñarse desde el inicio como una herramienta de acceso que preserve la privacidad, incorporando criterios de anonimización en la elaboración de los campos descriptivos.

Ejemplo de aplicación: *En lugar de publicar íntegramente un expediente de restitución de tierras que hace parte de un fondo histórico y que contiene datos sensibles de víctimas, puede publicarse una ficha descriptiva que indique el tipo de actuación, el período, la región geográfica general y el resultado del proceso, sin reproducir los nombres, relatos ni datos identificadores de las personas involucradas.*

Agregación estadística

La agregación estadística es una técnica de anonimización en la que la información sobre individuos se transforma en información sobre grupos. En lugar de divulgar datos asociables a personas

concretas, se divulgan totales, promedios, distribuciones, frecuencias o porcentajes calculados sobre el conjunto. Esta técnica es particularmente útil cuando la finalidad de la divulgación no requiere conocer información individual, sino patrones, tendencias o caracterizaciones del acervo documental.

En el contexto de la información no estructurada, la agregación puede aplicarse a través de mecanismos como:

i. Conteos sobre el universo documental, en los que se reportan cuántos expedientes presentan determinada característica sin identificar cuáles.
ii. Estadísticas descriptivas de menciones, frecuencias o coocurrencias temáticas en un corpus de textos.
iii. Resúmenes ejecutivos consolidados que reportan hallazgos generales sin atribución individual.
iv. Visualizaciones agregadas (gráficos, mapas de calor, líneas de tiempo) construidas a partir del corpus documental sin permitir el rastreo a documentos individuales.

La agregación es una técnica robusta cuando el tamaño del grupo es suficientemente grande como para impedir la singularización. Sin embargo, presenta riesgos específicos: cuando los grupos son pequeños o presentan combinaciones únicas de atributos, la agregación puede fallar y permitir la reidentificación por inferencia. Por esta razón, se recomienda aplicar reglas de tamaño mínimo de grupo (no reportar valores de grupos con menos de un umbral predefinido, comúnmente cinco o diez registros) y reglas de supresión secundaria (cuando la supresión de un valor permite deducirlo por diferencia con el total, suprimir también valores adicionales).

La agregación es complementaria, no sustitutiva, de las técnicas de intervención directa sobre el documento. Su uso resulta especialmente apropiado en escenarios de transparencia activa, publicación de datos abiertos, informes de gestión, estudios estadísticos sobre acervos documentales y reportes a organismos de control, cuando la finalidad puede satisfacerse sin acceso al documento individual.

Ejemplo de aplicación. *En lugar de publicar el listado de expedientes disciplinarios tramitados por una entidad en un período, puede publicarse un informe estadístico que reporte el número de procesos por causal, por tipo de falta y por dependencia, con reglas de tamaño mínimo de grupo que impidan la reidentificación por combinación de atributos. La publicación con datos*

individuales solo procederá cuando el universo poblacional sea suficientemente amplio y la combinación de atributos no permita la singularización.

Combinación de técnicas y criterios de selección

Ninguna de las técnicas descritas en los numerales anteriores es suficiente por sí sola para garantizar una anonimización efectiva en documentos no estructurados complejos. La práctica institucional y la literatura técnica especializada coinciden en que la anonimización robusta de este tipo de documentos requiere la aplicación combinada y estructurada de varias técnicas de manera simultánea, seleccionadas en función del análisis de riesgo previo.

La combinación adecuada depende de tres factores principales: el tipo de soporte del documento, el nivel de sensibilidad de los datos que contiene y la finalidad del acceso que se pretende habilitar. A manera de ejemplo, se presentan combinaciones orientadoras para los escenarios más frecuentes en archivos públicos colombianos:

- **Para expedientes judiciales o disciplinarios con múltiples intervinientes:** sustitución por etiquetas, más generalización de atributos contextuales, más supresión de datos sensibles no necesarios para la comprensión del proceso, más acceso mediado para partes protegidas
- **Para archivos del conflicto armado y derechos humanos:** disociación irreversible, más generalización geográfica y temporal, más separación de soportes, más resumen o abstracción para la versión de acceso público
- **Para grabaciones audiovisuales de audiencias o entrevistas:** distorsión técnica con seguimiento dinámico, más supresión de audio de fondo identificador, más, eliminación de metadatos, más acceso mediado presencial para versiones sin distorsión
- **Para publicación masiva de corpus documentales digitalizados:** tokenización semántica con revisión humana, más generalización
- **Para expedientes de salud, bienestar o protección social:** separación de soportes, más supresión de datos clínicos directos, más generalización de diagnósticos, más acceso mediado digital.

La selección y combinación de técnicas debe quedar documentada en el expediente de gestión del proceso de anonimización, indicando para cada técnica aplicada: el dato o elemento tratado, la técnica seleccionada, el fundamento de esa selección y el resultado del test de reidentificación posterior.

Técnicas insuficientes o no admisibles por sí solas

No toda intervención técnica que reduzca la visibilidad de un dato significa anonimización efectiva. Algunas técnicas, aplicadas de forma aislada, no neutralizan el riesgo de reidentificación y deben considerarse insuficientes. Su aplicación sin medidas complementarias puede comprometer la eficacia del proceso de anonimización. Se consideran técnicas insuficientes o no admisibles por sí solas, entre otras:

Tabla 5. Técnicas insuficientes o no admisibles por sí solas

Práctica	Por qué no funciona	Qué exige la anonimización
Tachado superficial en documentos PDF mediante anotaciones gráficas	Las herramientas que ocultan texto con un rectángulo encima dejan el dato subyacente recuperable mediante copia y pegado, búsqueda automática, OCR o inspección de capas	La redacción segura debe eliminar el dato del flujo de contenido del PDF, no solo cubrirlo visualmente
Difuminado o pixelado reversible	Los filtros de desenfoque o pixelado de baja intensidad pueden revertirse mediante mejora de imagen, super-resolución o fuerza bruta, sobre todo cuando el universo de referentes es reducido	Usar parámetros suficientemente robustos y, cuando proceda, técnicas no reversibles como la supresión absoluta del fragmento
Eliminación exclusiva del contenido visible sin tratamiento de capas técnicas	En PDF, Word, imágenes, audio y video, no atender metadatos embebidos, comentarios ocultos, control de cambios, miniaturas, capas OCR, propiedades del archivo o información del codificador deja vías de reidentificación abiertas	La intervención debe ser integral conforme al numeral 13.8
Renombramiento del archivo sin verificación de metadatos internos	Cambiar el nombre no modifica el contenido ni los metadatos embebidos: un archivo titulado “anonimizado.pdf” puede contener el nombre del autor original y el historial completo	Verificar e intervenir los metadatos internos además del nombre del archivo
Conversión simple de formato	Convertir, por ejemplo, de Word a PDF puede preservar la mayoría de los metadatos y, en algunos casos, agregar nuevos	La conversión no es por sí misma una técnica de anonimización; debe acompañarse de las técnicas correspondientes al formato resultante
Borrado lógico sin sobre escritura segura	Eliminar un archivo desde el sistema operativo o desde el sistema de gestión documental, no garantiza que la información no pueda recuperarse	La eliminación segura debe seguir las pautas del numeral 12.7
Sustitución de identificadores directos	Eliminar el nombre y número de identificación puede ser insuficiente si persisten	Analizar y tratar también los cuasi-identificadores

Práctica	Por qué no funciona	Qué exige la anonimización
sin atención a cuasi-identificadores	combinaciones de cuasi-identificadores (cargo, dependencia, fecha, hechos) que reducen el universo a una persona identificable	que permitan reducir el universo de personas posibles
Uso de capturas de pantalla como técnica de anonimización	La captura de un documento anonimizado mediante técnicas reversibles preserva la versión visible en imagen plana, pero si era recuperable, no agrega protección sustantiva. Además, puede incluir elementos del entorno del usuario, ajenos al documento original	No usar la captura como técnica de anonimización; aplicar técnicas irreversibles sobre el documento original

Fuente: elaboración propia

Estas técnicas pueden formar parte de un proceso de anonimización, solo cuando se combinen con medidas adicionales que neutralicen sus limitaciones. Su uso aislado no constituye anonimización efectiva y la entidad debe abstenerse de presentarlas como tales ante solicitudes de acceso, transferencias documentales o publicaciones.

Riesgos de reidentificación

La anonimización busca impedir que una persona pueda ser identificada a partir de un documento, ya sea de manera directa (porque su nombre aparece) o indirecta (porque la combinación de los datos restantes permite saber quién es). Cuando esa identificación sigue siendo posible, incluso después de aplicar una técnica de anonimización, se habla de riesgo de reidentificación.

Una anonimización deficiente puede llevar a que un tercero, con herramientas comúnmente disponibles, vuelva a poner nombre y apellido a una persona que el documento pretendía proteger. Las consecuencias afectan al titular cuya información protegida queda expuesta y a la entidad que asume las responsabilidades derivadas de una protección insuficiente.

La información puede estar adecuadamente anonimizada frente a un tercero externo y, al mismo tiempo, seguir siendo identificable para la propia entidad que conserva los expedientes originales. Esto no es un problema en sí mismo, debido a que la entidad necesita conservar el original para fines de gestión documental. Lo que debe garantizarse es que la versión que se da acceso, o publica, no permita la reidentificación con los medios razonablemente disponibles para quien la consulte.

Evaluación y registro del riesgo residual

El análisis de los riesgos de reidentificación no concluye con la aplicación de técnicas. La entidad debe evaluar el riesgo residual que persiste tras la intervención, utilizando las metodologías documentadas en el presente capítulo (prueba del intruso, modelo C-sanitized, análisis de los tres riesgos). Esta evaluación debe ser realizada por un equipo interdisciplinario (jurídico, archivístico y técnico) y sus resultados, junto con la justificación de la decisión, deben quedar registrados en el expediente del proceso de anonimización. La periodicidad de esta evaluación dependerá de la sensibilidad del documento y de los cambios en el contexto tecnológico que puedan afectar el riesgo de reidentificación.

Modalidades técnicas de riesgo de reidentificación

Los riesgos de reidentificación no se presentan de manera uniforme. Los textos técnicos identifican tres modalidades principales[19], cuya comprensión es indispensable para diseñar procesos de anonimización efectivos:

La reidentificación puede ocurrir por tres vías principales, que en la práctica suelen presentarse combinadas:

Tabla 6. Tres vías de reidentificación: singularización, vinculabilidad e inferencia

Singularización	Vinculabilidad	Inferencia
Aislar a una persona dentro del documento, aunque su nombre haya sido suprimido. **Ejemplo**: "la única funcionaria de la Subdirección de Talento Humano en comisión de servicios en julio de 2024".	Conectar el documento anonimizado con información disponible en otras fuentes. **Ejemplo:** "el contratista que ejecutó el convenio 045 de 2023" una búsqueda en SECOP lo recupera.	Deducir información sobre una persona a partir de los datos disponibles, sin que esté escrita. **Ejemplo:** "el servidor que se reincorporó tras una incapacidad prolongada en abril de 2024".

19 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

En la práctica, las tres vías se combinan: alguien singulariza al titular por la combinación de atributos del documento, lo vincula con información de redes sociales o registros públicos e infiere atributos sensibles que no estaban escritos.

Impacto de la inteligencia artificial sobre las modalidades de riesgo

El desarrollo acelerado de la inteligencia artificial presenta riesgos distintos a los tradicionalmente considerados en los procesos de anonimización. Los modelos de lenguaje de gran escala y 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 detectaría[20]. En el ámbito de los documentos de archivo, esto implica varios escenarios concretos:

a. **Texto narrativo.** Documentos anonimizados pueden ser procesados por modelos de lenguaje, para inferir la identidad del autor o del sujeto a partir del estilo redaccional, el vocabulario empleado o las referencias contextuales (Staab y otros, 2024).
b. **Imágenes.** Las imágenes con rostros distorsionados mediante técnicas reversibles, como pixelado o difuminado leve, pueden ser sometidas a procesos de mejora basados en inteligencia artificial que reconstruyan parcialmente los rasgos suprimidos (McPherson y otros, 2016). Las técnicas de enmascaramiento total o supresión completa de la región facial no presentan esta vulnerabilidad.
c. **Audio.** Los registros de voz con distorsión biométrica pueden ser procesados por modelos de separación de fuentes sonoras y de reconocimiento del hablante que recuperen características identificadoras[21].

20 Staab, R., Vero, M., Balunović, M., y 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

21 Wiepert, D. y otros. (2024). Reidentification of Participants in Shared Clinical Data Sets: Experimental Study. JMIR AI. Disponible en: https://ai.jmir.org/2024/1/e52054

d. **Datos disociados.** Conjuntos aparentemente disociados pueden ser correlacionados automáticamente con fuentes abiertas mediante técnicas de aprendizaje automático, que identifiquen coincidencias no evidentes para el análisis humano.

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 diseñadas para bases de datos estructuradas no fueron pensadas para enfrentar este tipo de análisis sobre texto libre (Lermen y otros, 2026)[22]. 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.

Los modelos de lenguaje de gran escala se entrenan con volúmenes masivos de información extraída de fuentes abiertas de internet. Información personal que fue publicada en algún momento, aunque posteriormente haya sido retirada o anonimizada, puede haber quedado parcialmente incorporada en los parámetros de estos modelos y, en determinadas condiciones, particularmente cuando se trata de secuencias largas, frecuentes o únicas en el corpus de entrenamiento, ser recuperable mediante técnicas de extracción (Carlini y otros, 2021; Carlini y otros, 2023)[23]. 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 publicación.

Múltiples publicaciones individualmente inocuas pueden, al cruzarse mediante técnicas de aprendizaje automático, conducir a la reidentificación de personas que en cada documento aislado habrían permanecido protegidas. Este efecto es especialmente relevante para entidades públicas, que por mandato de transparencia activa publican grandes volúmenes de documentos sobre los mismos sujetos, programas o territorios.

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.

---

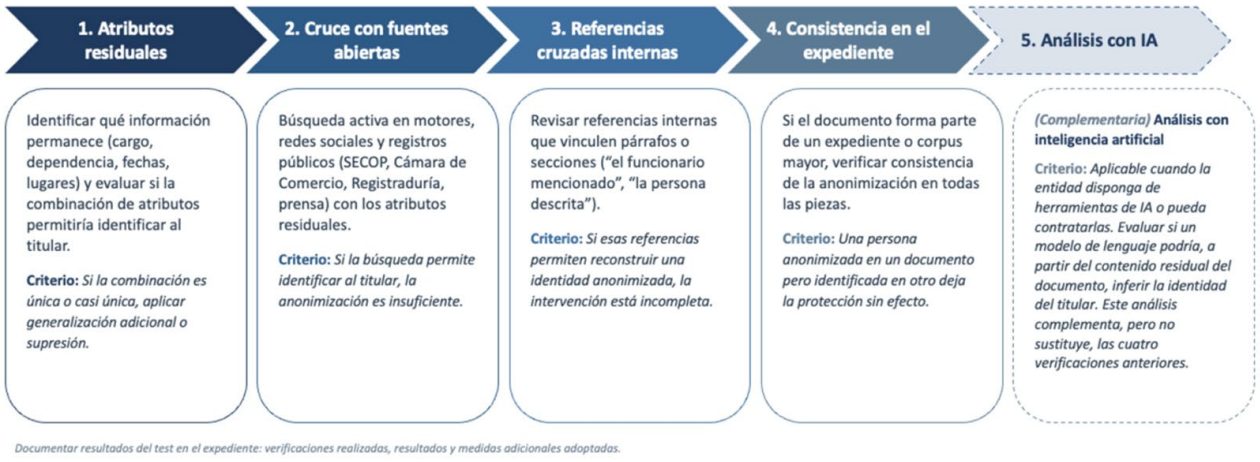
22 Lermen, S., Paleka, D., Swanson, J., Aerni, M., Carlini, N., y Tramèr, F. (2026). Large-scale online deanonymization with LLMs. arXiv preprint arXiv:2602.16800. Disponible en: https://arxiv.org/abs/2602.16800

23 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

Test de reidentificación

Antes de divulgar, transferir o publicar un documento anonimizado, la entidad debe realizar una verificación práctica que confirme si la intervención es suficiente. Este ejercicio se conoce como test de reidentificación. Es una herramienta de gestión que permite a la entidad demostrar, con evidencia documentada, que adoptó las medidas adecuadas. El test debe incluir, como mínimo, las cuatro verificaciones desarrolladas en el numeral 10.4 (Tabla 7) y aplicadas en el numeral 10.5 (Tabla 8). Tres de ellas son de aplicación general y no requieren herramientas especializadas, la prueba del intruso motivado, el modelo C-sanitized y el análisis de los tres riesgos; la cuarta, basada en pruebas de inferencia contextual mediante modelos de lenguaje, integra el uso de inteligencia artificial dentro del test, y se aplica cuando la entidad dispone de la herramienta o puede contratarla. Esta cuarta prueba refuerza la detección de riesgos no evidentes por técnicas estadísticas convencionales y opera de manera complementaria a las tres anteriores, sin sustituirlas.

Cuatro verificaciones obligatorias y una complementaria — Sección 10.3, Lineamientos AGN v2.0



Documentar resultados del test en el expediente: verificaciones realizadas, resultados y medidas adicionales adoptadas.

Los resultados del test deben quedar documentados en el expediente del proceso de anonimización: qué verificaciones se realizaron, qué resultados arrojaron y qué medidas adicionales se adoptaron cuando se detectaron riesgos. Esta documentación es parte del registro de trazabilidad y soporta la diligencia de la entidad ante eventuales requerimientos. Ninguna métrica elimina por completo el riesgo de reidentificación. La entidad debe documentar el nivel de riesgo residual asumido tras la intervención, los métodos aplicados para medirlo y las razones por las cuales se considera aceptable frente a la finalidad de la divulgación.

Métricas cuantitativas de referencia para evaluar la anonimización

En información no estructurada, las entidades deben acudir a métricas y pruebas diseñadas o adaptadas para texto libre, expedientes híbridos y registros audiovisuales:

Tabla 7. Métricas cuantitativas de referencia para evaluar la anonimización

Métrica	Aplicación y criterio
Prueba del intruso motivado (ICO, Anonymisation Code of Practice, 2012; actualización 2021)	Un evaluador con acceso a fuentes auxiliares razonables intenta reidentificar a los titulares en el documento sanitizado y se documenta la tasa de éxito. Es la métrica más operativa para *corpus* archivísticos y judiciales; no requiere infraestructura técnica compleja
Pruebas de inferencia contextual mediante modelos de lenguaje (Staab y otros, 2024)	La entidad somete el documento sanitizado a uno o varios modelos y mide la tasa de inferencia exitosa sobre atributos sensibles (ubicación, género, ocupación, condición de salud) deducibles a partir de combinaciones contextuales. Es la prueba más adecuada para detectar la reidentificación contextual, principal vector de riesgo en información no estructurada
Modelo C-sanitized (Sánchez y Batet, 2016)	Adaptación formal del k-anonimato a texto libre: mediante información mutua, evalúa si el riesgo de divulgación de una entidad protegida tras leer el documento se mantiene por debajo de un umbral definido. Aplica cuando la entidad cuenta con capacidad técnica para implementarlo o el *corpus* es suficientemente extenso para justificar el análisis automatizado
Análisis de los tres riesgos (GT29 / WP29, Dictamen 05/2014)	Criterio cualitativo transversal de cierre: singularización, vinculabilidad e inferencia. Aun cuando las pruebas anteriores arrojen resultados favorables, la entidad debe verificar que ninguno de los tres riesgos persista en el documento sanitizado

Fuente: elaboración propia

Las pruebas anteriores evalúan riesgo técnico de reidentificación. Adicionalmente, la entidad debe verificar la integridad contextual del flujo de información: si la divulgación, transferencia o publicación respeta la finalidad original de recolección y las expectativas legítimas del titular, conforme al principio de finalidad y al principio de circulación restringida. Un documento puede haber sido correctamente anonimizado y, aun así, su publicación puede vulnerar el contexto original de recolección cuando el cambio de receptor o de finalidad altera las condiciones bajo las cuales el titular entregó la información.

Cuando la entidad deba evaluar el riesgo de reidentificación sobre componentes estructurados del expediente (tablas, registros tabulares, microdatos), el marco operativo aplicable es el desarrollado en la Guía para la anonimización de datos estructurados del DANE (2024), que define el procedimiento de análisis de riesgos, identificación de pseudoidentificadores, selección de técnicas y evaluación de resultados aplicable a ese universo.

Advertencia sobre la prueba de inferencia mediante modelos de lenguaje.
La realización de la prueba de inferencia contextual a través de modelos de lenguaje, implica el envío del documento sanitizado a un sistema externo de procesamiento, lo cual constituye tratamiento de datos personales por parte de un encargado y, cuando el proveedor del modelo opera infraestructura ubicada fuera del territorio nacional, transferencia internacional de datos personales en los términos del artículo 26 de la Ley 1581 de 2012 y del artículo 2.2.2.25.5.1 del Decreto 1074 de 2015. La entidad debe considerar que el documento sometido a la prueba conserva por definición datos personales residuales bajo evaluación, razón por la cual su tratamiento por terceros no es jurídicamente equivalente al tratamiento de un documento ya anonimizado.

En consecuencia, antes de aplicar esta prueba, la entidad verificará que se cumpla al menos una de las siguientes condiciones: i) que la prueba se realice mediante modelos desplegados en infraestructura propia o contratada bajo control directo de la entidad, ya sea on-premise o en modalidad de despliegue dedicado con compromisos contractuales de no retención ni reutilización del contenido enviado; ii) que el proveedor del modelo cuente con cláusulas contractuales que garanticen un nivel adecuado de protección, o que opere desde un país declarado con el nivel adecuado por la Superintendencia de Industria y Comercio; o iii) que el documento sometido a la prueba haya sido previamente sometido a una intervención reforzada que elimine los identificadores directos y reduzca la singularización contextual, de modo que el riesgo residual sea proporcional al fin de verificación.

La entidad observará, en todo caso, los lineamientos sobre tratamiento de datos personales en sistemas de inteligencia artificial emitidos por la Superintendencia de Industria y Comercio. Cuando ninguna de las condiciones señaladas pueda satisfacerse, la entidad se abstendrá de aplicar esta prueba específica y soportará el test de reidentificación con las tres verificaciones restantes desarrolladas en la Tabla 7.

Las métricas de datos estructurados como referente conceptual

Los textos técnicos internacional han desarrollado métricas cuantitativas reconocidas para datos estructurados: k-anonimato, l-diversidad y t-cercanía, recogidas en marcos como ISO/IEC 20889 (2018), NIST IR 8053 (2015) y AEPD (2016). Estas métricas no son directamente aplicables a información no estructurada, porque están diseñadas para conjuntos con atributos discretos y registros homogéneos, condiciones que no se cumplen en textos narrativos, expedientes híbridos o registros audiovisuales, donde la singularización, la vinculabilidad y la inferencia derivan de patrones del lenguaje, referencias indirectas y coocurrencias contextuales. No obstante, conservan valor como referente conceptual y resultan aplicables a subconjuntos tabulables del corpus documental: por ejemplo, listados de personas con atributos categóricos extraídos del texto, anexos estadísticos o bases de datos asociadas a un expediente. En esos casos, la entidad puede aplicar k-anonimato y sus refinamientos al subconjunto estructurado, complementando la evaluación cualitativa del documento principal.

Balance riesgo-utilidad: el principio de proporcionalidad aplicado a la anonimización

La evaluación de la anonimización no se agota en la medición del riesgo de reidentificación. Conforme al principio de calidad de la información establecido en el artículo 3 de la Ley 1712 de 2014, y al principio de proporcionalidad, la entidad debe verificar, además del riesgo, si la versión que se va a dar acceso o publicar, conserva valor informativo suficiente para satisfacer la finalidad de la publicación o entrega y el derecho de acceso del solicitante. Una anonimización excesiva, es decir, aquella que suprime más de lo necesario y vacía el documento de contenido útil es una falla institucional simétrica a la anonimización insuficiente y puede comprometer el derecho fundamental de acceso a la información pública consagrado en el artículo 74 de la Constitución Política.

La utilidad informativa es la medida en que el documento sanitizado preserva el contenido necesario para cumplir la finalidad legítima que motiva su divulgación o entrega, depende tanto del propósito, como del destinatario.

Indicadores de utilidad para la información no estructurada. A diferencia de los datos estructurados, donde la utilidad se cuantifica con métricas algorítmicas sobre distribuciones, la información no estructurada (textos narrativos, expedientes híbridos, imágenes, audio y video) requiere indicadores cualitativos verificables aplicables al sentido del documento. La entidad debe evaluar, como mínimo:

i. preservación de las entidades funcionales del relato, qué ocurrió, dónde en términos generales, cuándo en escala temporal aceptable, con qué consecuencia institucional
ii. integridad narrativa, entendida como la posibilidad de comprender el sentido del documento sin el dato suprimido
iii. trazabilidad temporal suficiente para reconstruir la secuencia de hechos relevantes
iv. capacidad de respuesta efectiva a la pregunta del solicitante o a la finalidad declarada de la publicación. En soportes audiovisuales, estos indicadores se traducen en la conservación del contexto escenográfico, de los marcadores temporales y de los elementos no identificatorios, indispensables para comprender el hecho documentado.

La norma ISO/IEC 27559 (2022), Privacy enhancing data de-identification framework, aborda la de-identificación como un proceso que debe equilibrar riesgo de reidentificación y utilidad de los datos resultantes. En el mismo sentido se orientan las guías del ICO británico (Anonymisation: managing data protection risk, 2012; actualización 2021) y de la AEPD (Orientaciones y garantías en los procedimientos de anonimización de datos personales, 2016), que tratan la anonimización como una decisión de balance y no como una operación de supresión maximalista.

Aplicadas las cuatro pruebas del numeral 10.4 sobre la versión sanitizada, la entidad valora en paralelo la utilidad preservada y adopta una de cuatro posiciones:

a. riesgo aceptable y utilidad preservada, la versión puede publicarse o entregarse
b. riesgo aceptable, pero utilidad insuficiente, debe reconsiderarse la sanitización porque las supresiones no son proporcionales a la finalidad
c. riesgo no aceptable, pero utilidad recuperable, deben aplicarse técnicas adicionales de intervención (generalización, agregación, divulgación parcial motivada)
d. riesgo no aceptable y utilidad no recuperable mediante intervención, la divulgación no es viable en los términos solicitados y debe motivarse la denegación o la entrega bajo régimen de acceso restringido, conforme a los artículos 18 y 19 de la Ley 1712 de 2014.

El expediente del proceso de anonimización debe registrar no solo los riesgos identificados y las técnicas aplicadas, sino también la utilidad sacrificada en cada decisión: qué información se suprimió o generalizó, qué finalidad legítima dejó de servirse o quedó parcialmente atendida, y la justificación de proporcionalidad. Esta trazabilidad permite controlar la legalidad de la decisión tanto frente al titular como frente al solicitante de información pública, y constituye el soporte documental.

Caso aplicado

Una entidad del orden nacional cuya misionalidad es la inclusión social va a publicar en su portal web en la sección de "Transparencia", un informe trimestral de gestión sobre el programa de atención a víctimas del conflicto armado. La versión sanitizada eliminó nombres y números de identificación. El texto conserva el siguiente extracto:

> *"Durante el trimestre se atendieron 18 casos en el municipio X. Se destaca el caso de un líder social de 47 años, perteneciente a comunidad indígena, reubicado por amenazas tras denunciar actividades de minería ilegal en su territorio. La entidad gestionó subsidio de vivienda y acompañamiento psicosocial para él y su núcleo familiar de cinco integrantes. El beneficiario actualmente reside en zona urbana del departamento Y".*

Antes de publicar, la entidad aplica las cuatro pruebas del numeral 10.4:

Tabla 8. Pruebas de reidentificación aplicadas al caso del informe trimestral

Prueba	Preguntas guía	Resultado
Prueba del intruso motivado (ICO, Anonymisation Code of Practice, 2012; actualización, 2021)	• ¿Cuántos líderes sociales indígenas amenazados por denuncias de minería ilegal fueron registrados en el municipio X durante 2023? (consulta a la Defensoría del Pueblo, Sistema de Alertas Tempranas, prensa regional) • ¿La condición de líder reubicado permite cruzarlo con noticias de prensa local que cubrieron el caso? • ¿El RUV o el listado de beneficiarios del programa, accesibles por solicitud o publicados por otra dependencia, permiten ubicar a personas con ese perfil?	La combinación "líder indígena de 47 años, minería ilegal, municipio y reubicación reciente" suele singularizar a una sola persona en cualquier municipio del país. La prueba puede fallar si se hace una búsqueda en la prensa regional y probablemente devuelve el nombre en la primera consulta
Prueba de inferencia contextual con modelo de lenguaje (Staab y otros, 2024)	• ¿El modelo deduce la región específica del país más allá del municipio X? (si la combinación de minería ilegal, comunidad indígena y amenazas restringe a unas pocas regiones)	El modelo reconstruye perfil étnico, territorial, de seguridad y familiar (núcleo de cinco integrantes en zona urbana del departamento Y) sin necesidad del nombre. La prueba falla por inferencia

Prueba	Preguntas guía	Resultado
	• ¿Infiere el pueblo indígena específico? (puede deducirlo a partir del territorio y la dinámica minera descrita) • ¿Deriva atributos de seguridad? (el modelo identifica que se trata de una persona en situación de protección, lo que aumenta el riesgo si el documento llega a actores armados)	contextual, con un agravante, y es que el riesgo no es solo de privacidad, sino de seguridad personal
Modelo C-sanitized (cuando aplica) (Sánchez y Batet, 2016)	• ¿El documento sanitizado reduce la incertidumbre sobre la identidad del beneficiario por debajo del umbral C definido institucionalmente?	La densidad de cuasi-identificadores (etnia, edad, ocupación, motivo de amenaza, composición familiar y trayectoria) eleva la información mutua muy por encima del umbral. La prueba puede fallar
Análisis de los tres riesgos (GT29 / WP29, Dictamen 05 / 2014)	• **Singularización:** ¿el texto permite aislar a una persona dentro del universo posible? • **Vinculabilidad:** ¿permite vincular este registro con otros referidos a la misma persona? • **Inferencia:** ¿permite deducir atributos sensibles no declarados?	• Singularización: sí, la combinación de atributos lo identifica unívocamente en el municipio • Vinculabilidad: sí, cruza con prensa, RUV, registros de la UNP y denuncias previas. • Inferencia: sí, pertenencia étnica específica, ubicación protegida actual y riesgo de seguridad. Los tres riesgos persisten, con un componente adicional: la reidentificación no compromete solo la privacidad, sino la integridad física del beneficiario y su familia. La intervención no alcanza anonimización

Fuente: elaboración propia

Decisión institucional

A partir de los resultados de las cuatro pruebas, la entidad concluye que la versión sanitizada no alcanza el umbral de anonimización exigible y adopta medidas adicionales antes de publicar. La edad se generaliza a una formulación amplia (persona adulta); la pertenencia étnica específica se suprime o se reemplaza por una expresión genérica, y solo se conserva si resulta indispensable para la finalidad del informe; el motivo concreto de la amenaza se sustituye por una referencia general (víctima de hechos victimizantes en su territorio); la ubicación de origen y de reubicación se generalizan (un municipio del país, u otra región) o se eliminan cuando no sean esenciales, y la composición del núcleo familiar se omite. Adicionalmente, la entidad pondera si el caso debe figurar individualizado o si la finalidad del informe se cumple con datos agregados, por ejemplo, "18 casos atendidos, de los cuales X correspondieron a líderes sociales en situación de riesgo". Una vez aplicadas estas medidas, las cuatro pruebas se reaplican sobre la versión revisada y el nivel de riesgo residual, las pruebas practicadas y las decisiones adoptadas quedan documentados en el expediente del proceso de anonimización, como soporte de la responsabilidad demostrada de la entidad.

Consecuencias de una reidentificación

Una anonimización deficiente puede generar consecuencias serias, tanto para los titulares como para la entidad responsable. Las principales son:

a. **Consecuencias sobre los derechos fundamentales del titular.** La reidentificación no autorizada puede afectar el derecho al Habeas Data, el buen nombre, la autonomía personal y la autodeterminación informativa, reconocidos en el artículo 15 de la Constitución Política. En casos graves, puede vulnerar la dignidad humana, la igualdad y exponer al titular a escenarios de discriminación, difamación o estigmatización. La afectación es mayor cuando los datos reidentificados corresponden a categorías sensibles en los términos del artículo 5 de la Ley 1581 de 2012 (salud, orientación sexual, convicciones religiosas o políticas y datos biométricos, entre otros).

b. **Consecuencias económicas y laborales para el titular.** Una deficiente anonimización puede derivar en pérdida del empleo, reportes injustificados en bases de datos de morosos, perjuicios económicos por incumplimientos atribuidos indebidamente o restricciones en el acceso a servicios. Estos efectos son especialmente graves cuando los datos reidentificados corresponden a antecedentes judiciales, disciplinarios o de salud.

c. **Consecuencias físicas y psicológicas para el titular.** En contextos de conflicto armado, violencia de género, persecución política o pertenencia a grupos en situación de vulnerabilidad, la reidentificación puede contribuir al deterioro de la salud, al padecimiento psicológico, a la revictimización e incluso a la exposición a agresiones físicas o amenazas. En estos casos, el daño trasciende lo jurídico y compromete bienes irreversibles.

d. **Afectación a terceros relacionados.** La reidentificación de un titular puede exponer también a familiares, contactos, denunciantes, testigos o personas mencionadas en el mismo documento. En procesos judiciales, casos de violencia o expedientes sobre líderes sociales, la cadena de afectación puede extenderse a personas que no eran objeto principal del documento, pero quedaron vinculadas a él.

e. **Consecuencias institucionales para la entidad.** La entidad que aplica una anonimización deficiente asume distintos tipos de responsabilidad. No obstante, más allá de la responsabilidad jurídica, estos hechos erosionan la confianza ciudadana en las instituciones encargadas de custodiar la información pública y debilitan el ejercicio del derecho de acceso a la información.

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

Anonimización según el ciclo de vida documental

Consideración sobre niveles de intervención

La aplicación de técnicas de anonimización podrá variar en función del momento del ciclo de vida documental, el nivel de acceso requerido y el contexto de uso de la información. En términos generales, las entidades podrán adoptar enfoques diferenciados que incluyan:

- Medidas de control de acceso en etapas activas de gestión, cuando no se requiera intervención sobre el contenido
- Técnicas de seudonimización o protección parcial en fases de consulta interna o transferencia documental
- Procesos de anonimización irreversible en escenarios de divulgación pública o acceso abierto, especialmente en archivos históricos.

Estos niveles deberán definirse de manera coherente con los instrumentos archivísticos adoptados por la entidad.

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 de acceso, 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. Por eso 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 la 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 contemplar la preservación 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.

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.

Por ello, en el archivo de gestión la anonimización es 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 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).

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.

El archivo central debe estar listo para responder solicitudes de acceso. Esto no significa anonimizar todo documento que contenga datos personales: la regla general, fijada en la sección “Interpretación y aplicación”, es otra. La anonimización procede solo si la difusión íntegra del

documento, leída en su contexto, puede afectar derechos fundamentales o intereses constitucionalmente protegidos. La preparación, entonces, es preventiva en otro sentido: tener criterios, flujos y capacidades técnicas listos para usar si la situación llega. No es aplicarla por defecto. Cuando la anonimización no sea técnica o jurídicamente viable, la entidad puede recurrir a las medidas compensatorias y a los mecanismos de acceso controlado del Capítulo 15. Sobre esa base, se sugiere a la entidad aplicar 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 del 2000.

Sin perjuicio de lo anterior, la entidad deberá distinguir entre información pública clasificada e información pública 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. Un documento no podrá transferirse al archivo histórico mientras su reserva esté vigente conforme al artículo 22 de la Ley 1712 de 2014. Una vez

vencido el plazo legal de reserva, el documento pierde su condición de reservado y, si por sus condiciones intrínsecas tiene valor permanente, deberá transferirse al archivo histórico en los términos del numeral 11.4

- 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.

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.

La divulgación de documentos históricos no implica que la identidad de las personas deba hacerse pública sin filtros. La sola presencia de datos personales, incluso sensibles, no activa por sí misma la anonimización. La medida procede solo si la difusión íntegra del documento, en su contexto, afecta derechos fundamentales o intereses constitucionalmente protegidos sobre personas determinables, el ámbito propio de la información pública clasificada del artículo 18 de la Ley 1712 de 2014. Si ese es el caso, la entidad puede suprimir, disociar o limitar los elementos que identifiquen a los titulares. Cuando el documento llega al archivo histórico tras el vencimiento de una reserva del artículo 19 (defensa, seguridad, relaciones internacionales), la anonimización no rehabilita la denegación que regía durante la vigencia de la reserva, opera únicamente sobre los datos personales que conservan protección indefinida bajo el artículo 18 de la ley 1712 de 2014.

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 solo si, examinado el documento y su contexto de divulgación, se concluye que la difusión íntegra puede afectar derechos fundamentales o

intereses constitucionalmente protegidos. En los demás casos, el documento histórico se difunde íntegro, como manda la máxima publicidad

- Eliminar o suprimir identificadores directos (nombres, identificación, direcciones, firmas, imágenes u otros elementos) que permitan el reconocimiento inmediato, 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 un 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. Durante la vigencia de la reserva, el documento no puede transferirse al archivo histórico: su tránsito corresponde a las fases de gestión y central. Vencido el plazo legal de reserva conforme al artículo 22 de la Ley 1712 de 2014, el documento pierde su condición de reservado y, si concurre valor permanente, ingresa al archivo histórico, donde el acceso opera bajo la regla de máxima

publicidad sin perjuicio de la anonimización que proceda sobre los datos personales que conserven protección indefinida bajo el artículo 18.

De ahí que en el archivo histórico deba 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.

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.

Articulación con instrumentos archivísticos

La identificación, análisis y aplicación de técnicas de anonimización deberá articularse de manera obligatoria con los instrumentos archivísticos adoptados por la entidad, tales como los Cuadros de Clasificación Documental (CCD), las Tablas de Retención Documental (TRD), las Tablas de

Valoración Documental (TVD), las Tablas de Control de Acceso (TCA), los Programas de Gestión Documental (PGD) e inventarios documentales, en concordancia con el Acuerdo AGN 004 de 2019 y el Decreto 1080 de 2015.

En este sentido, las entidades podrán establecer criterios que permitan asociar niveles de sensibilidad de la información, condiciones de acceso y posibles técnicas de anonimización a las series, subseries o tipos documentales definidos en dichos instrumentos, con el fin de facilitar la toma de decisiones y promover la consistencia en el tratamiento de la información.

Pasos para la articulación de los regímenes aplicables

El reto institucional no radica en conocer cada régimen, protección de datos, transparencia y acceso a la información, gestión documental y regímenes especiales, sino en articularlos cuando un caso concreto los hace converger. El esquema que sigue ordena las preguntas que deben formularse, en secuencia, el funcionario o el equipo responsable de la decisión, e indica los criterios que activa cada paso. Se trata de una herramienta de orientación: no sustituye el análisis caso a caso ni releva a la entidad de su deber de motivación.

Paso 1. Identificación del escenario.
La entidad determina cuál es el supuesto que activa la decisión:

a. solicitud de acceso a la información (transparencia pasiva)
b. publicación proactiva, siempre y cuando no este exceptuado por el régimen de máxima publicidad;
c. uso interno con fines estadísticos, científicos o históricos
d. prueba tecnológica o entrenamiento de sistemas de inteligencia artificial
e. escenario distinto.

La tipificación del escenario determina los criterios aplicables.

Paso 2. Verificación del régimen aplicable (filtro de especialidad).
La entidad verifica, antes de aplicar el régimen general, si sobre el documento actúa un especial o si se encuentra excluido del ámbito de aplicación de la Ley 1581 de 2012.

a. **Exclusiones del ámbito de la Ley 1581 de 2012.** Si el documento corresponde a alguna de las hipótesis del artículo 2 de la Ley 1581 de 2012 (entre otras, bases para fines de seguridad y defensa nacional; inteligencia y contrainteligencia; información periodística;

bases reguladas por la Ley 1266 de 2008 sobre Habeas Data financiero; bases reguladas por la Ley 79 de 1993 sobre información estadística), no aplican las reglas de tratamiento del régimen general, sin perjuicio de la aplicación de los principios de protección de datos en los términos del parágrafo del mismo artículo.

b. **Regímenes especiales de reserva.** Si el documento se rige por un régimen especial que ordena reserva o tratamiento diferenciado, por ejemplo, Ley 1621 de 2013 (inteligencia y contrainteligencia), Ley 23 de 1981 (reserva médica), Ley 1448 de 2011 (víctimas), entre otros, las reglas particulares de ese régimen prevalecen y rigen la decisión, sin perjuicio de la articulación con los artículos 18 y 19 de la Ley 1712 de 2014 que corresponda según la naturaleza del interés tutelado.
c. **Inexistencia de régimen especial.** Si no se identifica exclusión ni régimen especial aplicable, se continúa con el análisis bajo el régimen general en los pasos siguientes.

En los supuestos de los literales a) y b), los pasos 3, 4 y 5 se ajustan o se sustituyen por las reglas del régimen especial, y la trazabilidad del paso 6 debe dejar constancia del fundamento normativo invocado.

Paso 3. Clasificación de acceso conforme a las TCA y la TRD

a. Si el documento, según lo establecido en las Tablas de Control de Acceso (TCA) y demás instrumentos archivísticos, está identificado como información pública, sin datos personales protegidos ni reserva, se da acceso o se publica de manera íntegra
b. Si está identificado como información clasificada (Ley 1712 de 2014, art. 18), reservada (Ley 1712 de 2014, art. 19) o sometida a régimen especial, continuar al paso 4.

Paso 4. Afectación a derechos fundamentales o intereses constitucionalmente protegidos.
La entidad valora si la divulgación íntegra del documento —esto es su entrega o publicación tal como reposa en el archivo, sin versión pública ni intervención técnica— puede generar alguno de los siguientes efectos:

a. Afectación a derechos de personas determinadas o determinables: vida, integridad personal, intimidad, Habeas Data, buen nombre, dignidad, prohibición de discriminación, derechos de niños, niñas y adolescentes o datos sensibles. Este es el ámbito propio de la información clasificada (Ley 1712 de 2014, art. 18).
b. Afectación a intereses públicos constitucional o legalmente protegidos: seguridad y defensa nacional, relaciones internacionales, debido proceso, prevención e investigación de delitos,

estabilidad macroeconómica, entre otros. Este es el ámbito propio de la información reservada (Ley 1712 de 2014, art. 19).

c. Afectación a regímenes especiales que ordenan reserva o tratamiento diferenciado (por ejemplo, Ley 1621 de 2013, Ley 23 de 1981, Ley 1448 de 2011).

- Si ninguno de estos efectos se configura, procede la difusión íntegra del documento
- Si se configura alguno, continuar al paso 5 para identificar la respuesta proporcional.

Paso 5. Respuesta proporcional.

La entidad valora las vías que ofrece el ordenamiento, en jerarquía de menor a mayor restricción del acceso:

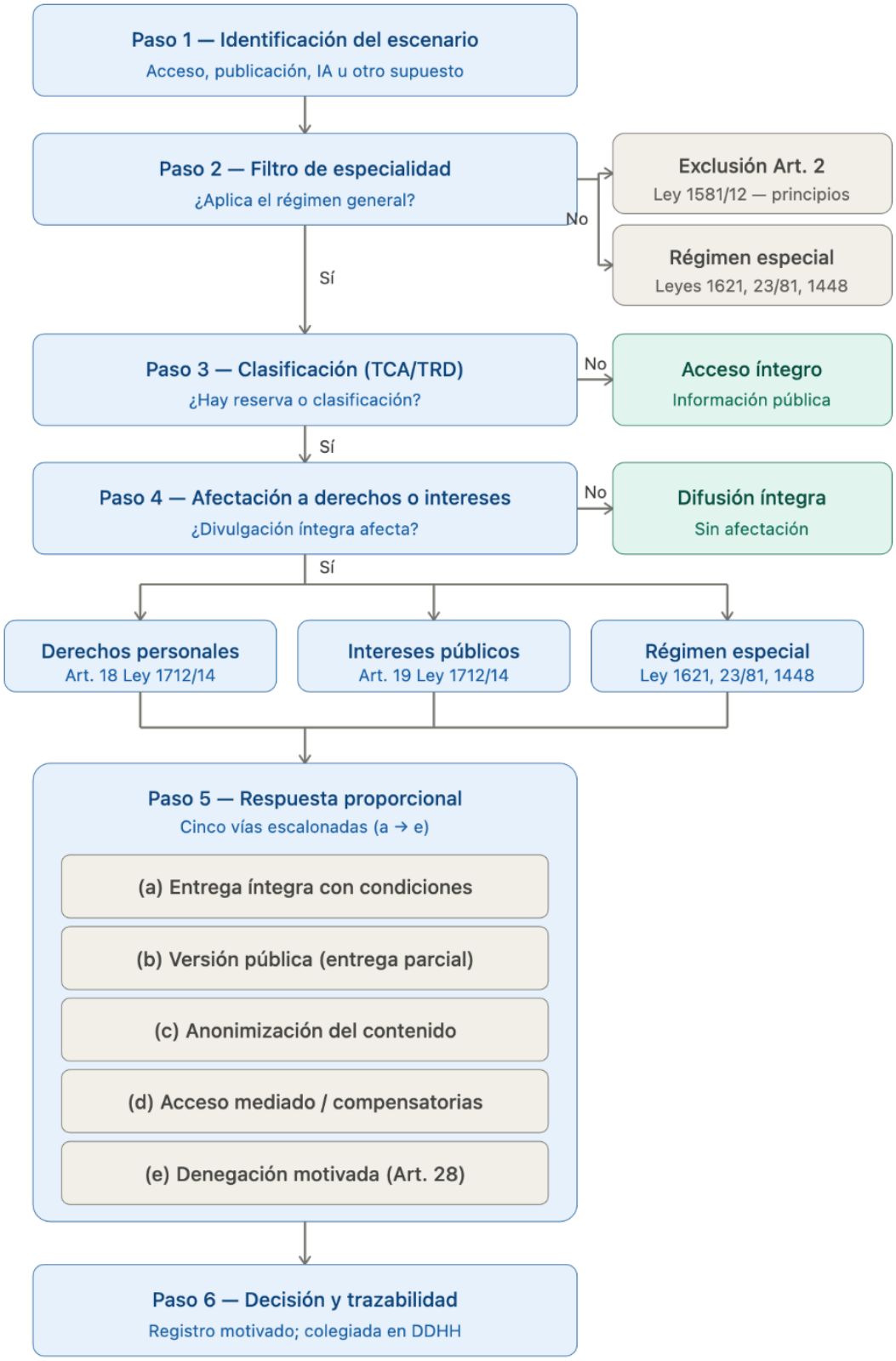
a. Entrega íntegra con condiciones de uso o compromiso de confidencialidad, cuando la afectación es marginal y la finalidad de la consulta lo justifica.
b. Entrega parcial mediante versión pública, cuando puede aislarse el contenido protegido sin perder el sentido del documento.
c. Anonimización del contenido protegido, cuando la entrega parcial no es viable o la finalidad requiere acceso al cuerpo del documento sin identificación de personas.
d. Acceso mediado o medidas compensatorias, cuando la anonimización no es técnicamente viable o el documento es indivisible.
e. Denegación motivada por clasificación o reserva (Ley 1712 de 2014, arículos 18 y 19), previa aplicación del test de daño. La denegación requiere acreditar, de manera concreta y motivada, que: i) el documento se subsume en una causal legal de excepción; ii) el daño que produciría la divulgación es presente, probable y específico; iii) el daño es mayor que el interés público en conocer la información, y iv) la restricción es proporcionada. El alcance del juicio de proporcionalidad varía según la categoría de excepción invocada. Para la información pública clasificada (artículo 18), la entidad acreditará haber descartado previamente todas las alternativas menos restrictivas de los literales a) al d), incluyendo la divulgación anonimizada. Para la información pública reservada (artículo 19), la anonimización ordinariamente no opera como alternativa a la denegación, dado que, suprimir identificadores personales no remueve por sí misma la afectación al interés público que motiva la reserva (defensa, seguridad nacional, relaciones internacionales, prevención e investigación de delitos, entre otros). No obstante, cuando del análisis del caso concreto se desprenda que el documento contiene segmentos cuyo contenido reservado es separable del contenido divulgable, de modo que la combinación de anonimización con redacción parcial o bifurcación del expediente permita habilitar el acceso a la información de interés público sin comprometer el bien jurídico tutelado por la reserva, la entidad valorará esa

alternativa como medida menos restrictiva antes de optar por la denegación, conforme al principio de máxima publicidad consagrado en los artículos 2 y 3 de la Ley 1712 de 2014. En todo caso, la entidad acreditará haber descartado las alternativas de los literales a), b) y d), y verificará que la reserva permanece dentro del plazo legal establecido.

Paso 6. Decisión y trazabilidad.

La decisión final queda formalmente registrada conforme al protocolo de trazabilidad institucional, con identificación del responsable, del régimen jurídico aplicado y de las técnicas utilizadas. Para casos relativos a archivos de derechos humanos, memoria histórica o eventos históricos de alto impacto, la decisión debe ser colegiada.

Flujo de decisión para la articulación de los regímenes aplicables



Fuente: elaboración propia

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
- **Reconocimiento de identificadores directos e indirectos:** la entidad deberá diferenciar entre: identificadores directos —que permiten el reconocimiento inmediato—, 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.

En el proceso de identificación de información susceptible de anonimización, las entidades podrán apoyarse en los instrumentos archivísticos, particularmente en las TRD, para reconocer los contextos funcionales, las series documentales y los tipos documentales en los que se produce la información.

Este análisis podrá complementarse con la identificación de los tipos de datos personales presentes en los documentos, de manera que se facilite la comprensión de la relación entre la estructura documental y los datos contenidos en ella, fortaleciendo la evaluación del riesgo y la selección de técnicas de anonimización.

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.

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 la fase del ciclo de vida documental:** la entidad deberá determinar si el documento pertenece a: un archivo de gestión (activo), archivo central o un 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 los datos personales, información sensible o a 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 del análisis jurídico
- **Verificación de información pública clasificada o pública reservada:** la entidad deberá determinar si el documento contiene información pública clasificada (artículo 18 de la Ley 1712 de 2014, que protege intereses privados como datos personales, secretos comerciales o derecho a la intimidad) o información pública reservada (artículo 19 de la Ley 1712 de 2014, que protege intereses públicos como defensa, seguridad nacional, relaciones internacionales o prevención de delitos). La distinción entre ambas categorías determina el tratamiento jurídico aplicable y debe quedar documentada en el expediente
- **En la información pública clasificada, la anonimización es un mecanismo válido de acceso parcial:** la entidad podrá anonimizar los identificadores que activan la protección privada y entregar el documento, siempre que la intervención preserve la inteligibilidad del documento. Esta es una de las hipótesis centrales de aplicación del presente lineamiento
- **En la información pública reservada, la anonimización resulta improcedente como mecanismo de acceso:** la afectación recae sobre el interés público mismo y no sobre la identidad de personas determinadas, razón por la cual suprimir nombres no remueve la causal de reserva. En estos supuestos la entidad deberá identificar la disposición legal que establece la reserva y determinar si procede la negación total o el levantamiento por vencimiento del término legal. La anonimización no podrá utilizarse como mecanismo para eludir una reserva legal vigente
- **Identificación de sujetos de especial protección:** la entidad deberá verificar si el documento involucra a 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:** con base 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.

En este punto, la entidad deberá diferenciar claramente entre los distintos regímenes que pueden converger en un mismo documento:

- **Reserva legal:** afecta intereses públicos (artículo 19 de la Ley 1712) y tiene un plazo definido. Durante su vigencia, la anonimización es improcedente
- **Clasificación:** protege derechos de personas determinables (artículo 18 de la Ley 1712) y es de carácter indefinido. La anonimización es un mecanismo válido para superar esta causal y permitir el acceso
- **Confidencialidad de datos personales:** protege la intimidad del titular, datos personales privados y sensibles (Ley 1581) y es indefinida, aplicando incluso después de la desclasificación
- **Valor histórico permanente:** no es una restricción, sino que fundamenta el acceso público a los documentos en el archivo histórico.

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.

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 a partir de medios razonables. Por lo tanto, la verificación debe: abarcar tanto el contenido principal como anexos, metadatos, elementos contextuales, capas técnicas ocultas y cualquier otro componente asociado. En expedientes híbridos, realizarse de manera transversal, asegurando coherencia entre soportes físicos, digitales y audiovisuales, y descartando la posibilidad de reconstruir la identidad por contraste entre ellos. De igual manera, la entidad debe 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 debe quedar formalmente validada por el área competente, dejando constancia interna del cumplimiento de las fases metodológicas y del control técnico final.

Prueba de no recuperación

Como parte de la verificación final, la entidad debe realizar pruebas activas que confirmen que el dato eliminado no puede recuperarse a partir de medios técnicos razonablemente disponibles. Estas pruebas comprenden, según el soporte:

- En documentos PDF, intentos de selección y copia del contenido aparentemente eliminado, búsqueda automática del texto suprimido e inspección del flujo de contenido del archivo
- En imágenes con desenfoque o pixelado, aplicación de filtros de mejora de imagen y verificación de que el resultado no permite reconstruir la información protegida
- En archivos de audio con distorsión de voz, aplicación de filtros inversos y verificación de que la voz no recupera reconocibilidad
- En documentos digitales con metadatos eliminados, inspección con herramientas especializadas que confirme la ausencia de metadatos identificables
- En versiones públicas derivadas de documentos con control de cambios o comentarios, verificación de que no quedan rastros de versiones anteriores
- La prueba de no recuperación debe quedar documentada con expresión de los métodos aplicados y los resultados obtenidos.

Revisión cruzada entre soportes

Cuando un mismo expediente o unidad documental conste en múltiples soportes físico y digital, copias maestras y copias administrativas, documento principal y anexos en formatos diversos, la verificación final debe incluir una revisión cruzada que confirme la coherencia de la anonimización a lo largo de todos los soportes. La incoherencia entre soportes, por ejemplo, anonimización de un nombre en la versión digital pero no en el escaneo del documento físico, o anonimización de un dato en el cuerpo del documento, pero no en un anexo vinculado compromete la totalidad del proceso, dado que la combinación de los soportes permite reconstruir la información protegida.

Registro formal de trazabilidad de la intervención

Toda intervención de anonimización debe quedar consignada en un registro institucional formal, acta, ficha técnica, asiento en el sistema de gestión documental electrónica o instrumento equivalente, que documente, al menos:

- Identificación del documento o expediente intervenido
- Régimen jurídico que sustenta la intervención
- Técnicas aplicadas
- Componentes intervenidos (cuerpo, anexos, metadatos, capas técnicas)
- Responsable técnico de la intervención
- Responsable institucional que valida el resultado
- Resultados de la prueba de no recuperación
- Versión a la que corresponde la intervención (versión pública, copia administrativa, copia para transferencia, entre otras)
- Fecha y firma o validación electrónica equivalente.

Este registro forma parte del expediente y constituye soporte probatorio de la diligencia de la entidad frente a los titulares, las autoridades de control y los terceros con interés legítimo. La omisión de la verificación final compromete la eficacia de la anonimización y puede generar responsabilidad por divulgación indebida de información protegida.

Eliminación segura de datos en procesos de anonimización

Los procesos de anonimización generan, en distintas fases, archivos y soportes que contienen datos personales no anonimizados o parcialmente anonimizados: copias de trabajo, versiones intermedias, archivos temporales del sistema, copias de respaldo y, en algunos casos, soportes

tecnológicos completos al final de su vida útil. La sola aplicación de técnicas de anonimización sobre el documento de divulgación no garantiza por sí misma la protección de los titulares si los archivos intermedios y los soportes tecnológicos no son objeto de eliminación segura cuando dejan de ser necesarios.

Distinción entre eliminación lógica y eliminación segura. La eliminación lógica, el borrado convencional desde el sistema operativo, no garantiza la imposibilidad de recuperación, dado que los datos permanecen en el medio de almacenamiento hasta que son sobrescritos. La eliminación segura, en cambio, garantiza que la información no pueda ser recuperada por medios técnicos razonablemente disponibles.

Técnicas aplicables según el soporte. Las entidades pueden aplicar, según el soporte:

- **Soportes físicos (papel, microfilm, fotografía análoga):** destrucción física certificada por trituración, incineración o pulverización, conforme a la Ley 594 de 2000 y con las garantías documentales aplicables (acta de eliminación, certificado del prestador del servicio cuando se contrata externamente)
- **Soportes digitales magnéticos (discos duros mecánicos, cintas):** sobrescritura múltiple conforme a estándares como NIST SP 800-88 Rev. 1 (2014) o, alternativamente, desmagnetización con equipos certificados. La sobrescritura simple no es suficiente para soportes con datos sensibles
- **Soportes digitales de estado sólido (SSD, memorias flash, dispositivos móviles):** eliminación mediante comandos específicos del firmware (ATA Secure Erase, NVMe Format con Crypto Erase) o destrucción física, dado que la sobrescritura tradicional resulta menos efectiva en estos medios por su arquitectura interna
- **Eliminación criptográfica:** cuando los datos están almacenados con cifrado robusto, la eliminación segura puede consistir en la destrucción irrecuperable de la clave, lo que vuelve los datos cifrados inaccesibles. Esta técnica es especialmente útil en entornos de nube y servicios distribuidos
- **Trazabilidad:** toda eliminación segura debe quedar documentada con expresión, al menos, del soporte o archivo eliminado, la fecha, el responsable, la técnica aplicada, el motivo y la autorización institucional cuando aplique. La documentación de la eliminación y constituye soporte probatorio frente a eventuales requerimientos de las autoridades de control o autoridades judiciales
- **Coherencia con TRD y TVD:** la eliminación segura no autoriza a la entidad para apartarse de las decisiones de retención y disposición final, establecidas en las TRD y TVD. Las copias de trabajo, versiones intermedias y archivos temporales generados durante el proceso de

anonimización pueden eliminarse, cuando su finalidad operativa haya cesado y siempre que su existencia no esté prevista en los instrumentos archivísticos como elemento integrante del expediente.

Cuadro resumen del proceso metodológico

Con el fin de facilitar la comprensión integral del proceso descrito en el presente capítulo, se incorpora a continuación un instrumento de síntesis que consolida sus etapas. Este instrumento no sustituye el desarrollo normativo contenido en los numerales precedentes, ni introduce obligaciones adicionales; cumple una función orientadora para reforzar la trazabilidad, la proporcionalidad de la intervención y la articulación entre las fases jurídicas y técnicas del proceso.

Tabla 9. Cuadro resumen del proceso metodológico de anonimización

Fase	Numeral	Propósito	Resultado esperado
Identificación	12.3	Determinar datos susceptibles de anonimización	Documento analizado
Régimen jurídico	12.4	Definir habilitación legal y límites	Concepto jurídico
Intervención técnica	13.1 a 13.8	Aplicar la técnica según el formato	Versión intervenida
Verificación final	12.6	Validar integridad y riesgo residual	Documento validado
Alternativas	15.1 y 15.2	Activar medidas compensatorias o de acceso controlado cuando aplique	Acceso controlado o decisión motivada
Eliminación segura	12.7	Eliminar copias intermedias y soportes que dejaron de ser necesarios	Soportes eliminados con trazabilidad

Fuente: elaboración propia

Intervención técnica según soporte documental

Reglas generales

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 a partir de 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 desde 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 la 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 la recuperación técnica del contenido o la reidentificación por correlación indirecta.

En todos los casos, la intervención técnica deberá respetar los principios archivísticos de procedencia, orden original e integridad documental. La anonimización opera sobre copias de consulta o versiones públicas y, por lo tanto, no debe afectar la estructura intelectual, la organización ni el contexto archivístico del expediente original, garantizando que la versión pública conserve su sentido y coherencia como unidad documental.

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 una 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 la 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.

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. Por eso la anonimización en formato

digital no puede 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.

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 un 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 la recuperación razonable del contenido eliminado ni la 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.

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 recomienda 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

- Como alternativa a la distorsión vocal en posproducción, podrá utilizarse la técnica de relectura por narrador distinto, cuando el testimonio sustantivo deba conservarse, pero la voz del titular comprometa el anonimato. La transcripción del relato debe respetar la voz narrativa original y la versión leída debe identificarse expresamente como una recreación.

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 la 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
- En proyectos audiovisuales con finalidad de divulgación, memoria o pedagogía, la dramatización con actores, la reconstrucción en locaciones neutras y la reinterpretación narrativa constituyen alternativas válidas a la distorsión técnica de imagen y sonido, siempre que se documente expresamente la naturaleza derivada del material y se obtengan las autorizaciones de uso correspondientes.

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. Por lo tanto, se sugiere que:

- El expediente se entienda como una unidad documental integral. Así, 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.

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.

Selección de técnicas y análisis de riesgo contextual

La selección de la técnica de anonimización no es una decisión discrecional ni uniforme. Depende del soporte documental, del nivel de riesgo de reidentificación y del entorno de divulgación previsto. El presente capítulo establece los lineamientos institucionales para articular esos tres factores en una decisión técnica trazable y proporcional, en operacionalización del marco conceptual desarrollado en la primera parte de estos lineamientos.

Lineamientos para la selección de técnicas

Las reglas técnicas de intervención por soporte documental se desarrollan en el numeral 13 del presente documento. Este capítulo articula tales reglas con el análisis de riesgo contextual del numeral 14.2 y con la matriz de decisión del numeral 14.3, de modo que la selección de la técnica no obedezca únicamente al soporte, sino al cruce entre soporte, nivel de riesgo identificado y entorno de divulgación previsto.

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 será efectiva si se limita a ocultaciones superficiales que permitan la recuperación técnica del contenido o la reidentificación por correlación indirecta, ni tampoco cuando, por exceso de supresión, vacíe al documento del contenido necesario para cumplir la finalidad legítima de su divulgación, en desmedro de los principios de máxima publicidad y calidad de la información (Ley 1712 de 2014).

Graduación de la técnica según el nivel de riesgo

La intensidad de la técnica aplicada deberá graduarse según el nivel de riesgo identificado conforme al numeral 14.2:

a. En escenarios de riesgo bajo, podrá resultar suficiente la supresión o sustitución básica acompañada de revisión técnica estándar
b. 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
c. 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 la recuperación técnica del contenido o la reidentificación por correlación indirecta.

Por otro lado, tampoco será adecuada cuando, por exceso de supresión, vacíe al documento del contenido necesario para cumplir la finalidad legítima de su divulgación o entrega y desconozca los principios de máxima publicidad y calidad de la información, en el marco de la Ley 1712 de 2014. La idoneidad de la técnica se mide por el balance del riesgo-utilidad evaluado conforme al numeral 10.5.

Análisis de riesgo contextual aplicado a la selección de técnicas

El capítulo 10 de la primera parte desarrolló el marco conceptual del riesgo de la reidentificación: las modalidades técnicas (singularización, vinculabilidad e inferencia), el impacto de la inteligencia artificial sobre estas modalidades, el test de reidentificación, las métricas cuantitativas aplicables a información no estructurada (prueba del intruso motivado, pruebas de inferencia con modelos de lenguaje, modelo C-sanitized y análisis de los tres riesgos del Dictamen 05/2014) y las consecuencias institucionales de una reidentificación deficiente.

Así las cosas, la función del presente numeral es operativa, al traducir el resultado del análisis realizado conforme al capítulo 10 en una clasificación funcional de bajo, medio o alto, que orienta la selección de la técnica conforme a la matriz del numeral 14.3 y, en su caso, la activación de las

medidas alternativas del capítulo 15. El análisis de riesgo contextual constituye una fase obligatoria del proceso institucional de anonimización. Su realización corresponde a las pruebas establecidas en el capítulo 10. Sus resultados deberán quedar documentados en el expediente del proceso de anonimización.

Clasificación funcional del riesgo

A partir del análisis realizado conforme al capítulo 10, la entidad clasificará el riesgo en uno de los siguientes tres niveles:

a. Riesgo bajo. La identificación del titular requiere esfuerzos desproporcionados o acceso a información no razonablemente disponible. Las pruebas de reidentificación arrojan resultados favorables y los tres riesgos del Dictamen 05/2014 no persisten de manera significativa.
b. Riesgo medio. Existe posibilidad razonable de identificación mediante combinación de fuentes accesibles. Al menos una de las pruebas del capítulo 10 detecta vectores de reidentificación que requieren intervención adicional, sin que la afectación sea de tal magnitud que impida la divulgación con técnica reforzada.
c. Riesgo alto. La identificación es probable con información pública o fácilmente inferible. Las pruebas del capítulo 10 detectan reidentificación efectiva en uno o varios vectores; la combinación de cuasi-identificadores singulariza al titular o al contexto narrativo y permite inferencia directa de atributos sensibles.

Consecuencias operativas de la clasificación

La clasificación del riesgo determina la respuesta institucional:

a. Cuando el riesgo es bajo, la entidad podrá proceder con la técnica base correspondiente al soporte conforme a la matriz del numeral 14.3, sin medidas adicionales más allá de la verificación final del numeral 12.6.
b. Cuando el riesgo es medio, la entidad reforzará la anonimización mediante generalización adicional, supresión contextual o combinación de técnicas conforme al numeral 14.3, y reaplicará el test de reidentificación sobre la versión revisada.
c. Cuando el riesgo es alto, la entidad optará por combinación intensa de técnicas, restricción del acceso, divulgación parcial debidamente motivada o, cuando proceda, activación de las medidas alternativas previstas en el capítulo 15. La divulgación íntegra del documento queda excluida mientras subsista el nivel de riesgo.

Requisito de efectividad

La anonimización se considerará adecuada únicamente cuando, tras la aplicación de la técnica seleccionada, la probabilidad de reidentificación a través de 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.

La omisión del análisis de riesgo contextual o la aplicación de técnicas no proporcionales al nivel de riesgo identificado, convierte la intervención en una supresión formal de datos sin efecto anonimizador real, y compromete la diligencia institucional frente a las autoridades de control.

Matriz de decisión: soporte, nivel de riesgo y técnicas recomendadas

La siguiente matriz articula los lineamientos del numeral 14.1 con la clasificación funcional del riesgo del numeral 14.2. Es una guía indicativa: las particularidades del documento pueden requerir combinaciones distintas. La matriz no sustituye el análisis de riesgo contextual realizado conforme al capítulo 10, ni la verificación final del numeral 12.6; por su parte, sí ofrece un marco de referencia para anclar la decisión técnica en parámetros estandarizados. La entidad debe documentar la combinación seleccionada y su justificación.

Tabla 10. Matriz de decisión: soporte, nivel de riesgo y técnicas recomendadas

Soporte / Formato	Riesgo bajo	Riesgo medio	Riesgo alto
Documento físico **(papel)**	Supresión local con tachado seguro; generalización de cuasi-identificadores cuando se reproduzca digitalmente	Supresión + generalización + control institucional de copias y reproducciones; bifurcación documental cuando aplique	Separación de soportes + acceso mediado; copia anonimizada para consulta y original con acceso restringido
Texto digital **(Word, PDF, correos)**	Supresión segura del contenido + revisión y eliminación de metadatos embebidos	Supresión + generalización + eliminación de metadatos + revisión de capas técnicas no visibles (numeral 13.8)	Combinación de técnicas + verificación cruzada de capas + agregación o resumen cuando proceda
Imágenes **(JPEG, TIFF, PNG)**	Pixelado o desenfoque de elementos identificables + eliminación de metadatos EXIF	Desenfoque robusto + supresión EXIF + revisión de miniaturas embebidas y de capas en imágenes editadas	Combinación de técnicas + verificación de imposibilidad de recuperación por mejora técnica + reformulación cuando aplique

Soporte / Formato	Riesgo bajo	Riesgo medio	Riesgo alto
Audio	Distorsión técnica básica de identificadores audibles	Distorsión de voz + supresión de fragmentos contextuales identificables + generalización narrativa	Desidentificación completa + revisión segmento a segmento + supresión de metadatos del codificador
Video	Distorsión técnica básica de rostros y elementos visibles	Desenfoque selectivo con seguimiento dinámico + distorsión de voz + supresión de secuencias contextuales identificables	Desidentificación completa + revisión cuadro a cuadro + supresión de metadatos del codificador
Documentos escaneados **(mixto)**	Supresión sobre capa OCR + supresión visual del contenido + verificación de coincidencia entre ambas capas	Supresión + generalización + eliminación de metadatos + verificación de elementos accesorios visibles (sellos, códigos QR, firmas, huellas, marcas manuscritas, números de radicación)	Separación de soportes o reformulación del contenido; reproducción segura como imagen plana sin OCR cuando proceda
Expedientes híbridos y multimodales	Aplicación coherente de técnicas en cada componente del expediente, con codificación uniforme cuando se use sustitución	Combinación de técnicas por componente + verificación de coherencia transversal + atención a anexos vinculados	Bifurcación del expediente + acceso mediado por canal + revisión cruzada exhaustiva entre componentes

Fuente: elaboración propia

Las técnicas referenciadas en la matriz se desarrollan en el capítulo de técnicas de la primera parte y los lineamientos de intervención por soporte se encuentran en el capítulo 13 de la presente segunda parte.

Medidas alternativas y mecanismos de acceso controlado

Medidas compensatorias cuando la anonimización no sea viable

Aun cuando la anonimización puede utilizarse como mecanismo técnico para armonizar el derecho de acceso a la información pública con la protección de datos personales, existen escenarios en los

que su aplicación resulta materialmente inviable, técnicamente desproporcionada o jurídicamente improcedente. En estos casos, la entidad puede acudir a medidas compensatorias que reduzcan el riesgo de identificación y permitan cumplir el marco legal de forma proporcional, sin que ello implique una renuncia a las obligaciones de protección.

Criterios de procedencia. La adopción de medidas compensatorias debe responder a los siguientes criterios concurrentes:

- **Proporcionalidad** entre el medio escogido y el fin perseguido (acceso, transparencia, conservación, finalidad histórica o investigativa, según corresponda)
- **Finalidad legítima** de la consulta o divulgación, y existencia de una base jurídica habilitante
- **Control institucional efectivo** sobre el acceso, documentado y verificable.

La medida compensatoria es una decisión motivada de la entidad, no una opción discrecional, y debe quedar registrada en el expediente correspondiente.

Catálogo de medidas compensatorias. Cuando la anonimización no sea viable, la entidad podrá adoptar una o varias de las siguientes medidas, según el tipo de soporte, el nivel de riesgo y la finalidad del acceso:

a) **Consulta en entorno físico controlado.** Acceso a los documentos únicamente en sala designada de la entidad, con supervisión por personal autorizado, restricción de retiro de documentos del área autorizada y control sobre dispositivos de captura del usuario.
b) **Restricción de dispositivos de captura y reproducción.** Prohibición o limitación del uso de teléfonos móviles, cámaras fotográficas o de video, escáneres portátiles y dispositivos de almacenamiento externo durante la consulta.
c) **Consulta en entornos digitales controlados.** Cuando la consulta se realice en formato digital, 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, perfiles de acceso con permisos limitados e implementación de entornos de consulta supervisados.
d) **Registro y trazabilidad del acceso.** Mecanismos que permitan identificar al usuario, registrar la fecha y hora del acceso, los documentos consultados y la finalidad de la consulta cuando corresponda.
e) **Acceso supervisado.** La consulta de determinados documentos se realiza únicamente en presencia de un funcionario designado, particularmente cuando se trate de documentos con datos sensibles o con valor probatorio.

f) **Acceso por extractos o transcripciones autorizadas.** Entrega de fragmentos seleccionados o transcripciones supervisadas que sustituyen la consulta directa al documento, conservando la posibilidad de verificación posterior.
g) **Mediación profesional o institucional.** Intervención de un profesional designado por la entidad o de una instancia colegiada que medie entre el solicitante y el documento, especialmente cuando se trate de archivos de derechos humanos, de inteligencia o de memoria histórica.

Reglas reforzadas para datos sensibles e información reservada. Cuando los documentos contengan datos sensibles, en los términos del artículo 5 de la Ley 1581 de 2012 y el artículo 3 del Decreto 1377 de 2013, o información clasificada o reservada conforme a los artículos 18 y 19 de la Ley 1712 de 2014, las medidas compensatorias deben reforzarse mediante:

- Restricción máxima del universo de personas con acceso
- Supervisión continua durante la consulta
- Compromiso documental de no divulgación
- Trazabilidad reforzada con registro detallado del acceso
- Cuando aplique, autorización expresa de la autoridad competente.

La medida compensatoria no autoriza a la entidad a apartarse de las garantías sustantivas del régimen aplicable, sino que las opera por una vía distinta a la anonimización. La implementación de estas medidas no exime a la entidad del deber de aplicar la anonimización en los demás documentos del mismo acervo en los que sí resulte técnicamente viable, ni del deber de revisar periódicamente la procedencia de la medida compensatoria conforme a la evolución de los recursos técnicos disponibles.

Mecanismos institucionales de acceso controlado

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. Entre estas medidas se incluyen las siguientes:

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 TRD, 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.

Estructura institucional y responsabilidades

Configuración institucional según el escenario

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 especialmente:

- El área productora del documento
- La Oficina Jurídica o quien ejerza el análisis del régimen aplicable
- Gestión Documental, cuando implique la 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.

Automatización y apoyo tecnológico: las entidades podrán apoyarse en herramientas tecnológicas que faciliten la identificación, análisis y aplicación de técnicas de anonimización sobre información no estructurada, incluyendo soluciones basadas en reconocimiento de texto, procesamiento de lenguaje natural, análisis de imágenes y otras tecnologías disponibles.

En escenarios de alto volumen o complejidad, se recomienda evaluar la implementación de mecanismos que permitan la aplicación consistente y trazable de las técnicas de anonimización, garantizando la integridad del documento original y la generación de versiones derivadas conforme a los lineamientos establecidos.

En entornos digitales, la información no estructurada puede originarse o gestionarse a través de sistemas de información, por lo que las entidades podrán evaluar mecanismos que permitan integrar los procesos de anonimización con dichos sistemas, garantizando consistencia en el tratamiento de la información, trazabilidad de las intervenciones realizadas y control sobre las versiones derivadas para acceso o divulgación.

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.

Por ello, 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 la comprensión pública.

La anonimización no puede vaciar el contenido sustancial ni alterar el sentido del documento. De ahí que el área misional deba 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

Tabla 11. Secuencia metodológica y responsabilidades institucionales

Fase	Etapa 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	Oficina Jurídica	Área productora	Acta de análisis

Fase	Etapa metodológica	Alcance por tipo de información no estructurada	Dependencia líder	Dependencias que intervienen	Producto o evidencia
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, edición, sustitución, etc.)	Dependencia designada	TI	Versión intervenida
7	Eliminación de metadatos y verificación técnica	Digital, imagen, audio, video	TI	Gestión Documental	Registro técnico
8	Revisión cruzada jurídica y técnica	Todos los soportes	Oficina Jurídica	Gestión Documental	Validación formal
9	Evaluación de riesgo residual	Todos los soportes	Oficina 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

BIBLIOGRAFÍA

Normativa colombiana

Constitución Política

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

Leyes

Colombia. Congreso de la República. (15 de enero de 1972). Ley 13 de 1972, sobre voto secreto. *Diario Oficial* No. 33.752.

—. (18 de febrero de 1981). Ley 23 de 1981, por la cual se dictan normas en materia de ética médica. Diario Oficial No. 35.711. [Artículos 33 y ss.: Reserva de la Historia Clínica].

—. (18 de agosto de 1989). Ley 80 de 1989, por la cual se crea el Archivo General de la Nación. *Diario Oficial* No. 38.853.

—. (20 de octubre de 1993). Ley 79 de 1993, por la cual se regula la realización de los Censos de Población y Vivienda en todo el territorio nacional. *Diario Oficial* No. 41.082.

—. (1996). Ley 256 de 1996, por la cual se dictan normas sobre competencia desleal. [Artículo 16: Secretos empresariales].

—. (14 de julio del 2000). Ley 594 de 2000, Ley General de Archivos. *Diario Oficial* No. 44.084.

—. (23 de enero del 2006). Ley 1010 de 2006, por medio de la cual se adoptan medidas para prevenir, corregir y sancionar el acoso laboral y otros hostigamientos en el marco de las relaciones de trabajo. *Diario Oficial* No. 46.160. [Artículo 14: Reserva de las actuaciones].

—. (8 de noviembre del 2006). Ley 1098 de 2006, Código de la Infancia y la Adolescencia. *Diario Oficial* No. 46.446. [Artículo 47].

—. (10 de junio del 2011). Ley 1448 de 2011, Ley de Víctimas y Restitución de Tierras. *Diario Oficial* No. 48.096.

—. (31 de diciembre del 2008). Ley 1266 de 2008, por la cual se dictan las disposiciones generales del hábeas data y se regula el manejo de la información contenida en bases de datos personales, en especial la financiera, crediticia, comercial, de servicios y la proveniente de terceros países. *Diario Oficial* No. 47.219. [Artículo 3, literales g) y h): definiciones de dato semiprivado y dato privado].

—. (18 de enero del 2011). Ley 1437 de 2011, Código de Procedimiento Administrativo y de lo Contencioso Administrativo. *Diario Oficial* No. 47.956. [Artículo 24].

—. (17 de octubre del 2012). Ley 1581 de 2012, por la cual se dictan disposiciones generales para la protección de datos personales. *Diario Oficial* No. 48.587. [Artículos 4, 5, 6 y 10].

—. (17 de abril del 2013). Ley 1621 de 2013, normas sobre actividades de inteligencia y contrainteligencia. *Diario Oficial* No. 48.764. [Artículo 33].

—. (6 de marzo del 2013). Ley 1618 de 2013, disposiciones para el pleno ejercicio de los derechos de las personas con discapacidad. *Diario Oficial* No. 48.717.

—. (6 de marzo del 2014). Ley 1712 de 2014, Ley de Transparencia y del Derecho de Acceso a la Información Pública Nacional. Diario Oficial No. 49.084. [Artículos 2, 18, 19 y 21].

—. (2023). Ley 2335 de 2023, disposiciones sobre las estadísticas oficiales en el país. *Diario Oficial*.

—. (1986). Código Electoral. [Artículo 51].

Colombia. Estatuto Tributario. [Artículo 583: Reserva de la información tributaria].

Decretos

Colombia. Presidencia de la República. (27 de junio del 2013). Decreto 1377 de 2013, por el cual se reglamenta parcialmente la Ley 1581 de 2012. *Diario Oficial* No. 48.834. [Artículo 26: Transferencia internacional de datos personales, compilado en el Decreto 1074 de 2015].

—. (21 de marzo del 2014, 21 de marzo). Decreto 857 de 2014, por el cual se reglamentan artículos de la Ley 1621 de 2013. *Diario Oficial* No. 49.102.

—. (26 de mayo del 2015a). Decreto 1070 de 2015, Decreto Único Reglamentario del Sector Administrativo de Defensa. *Diario Oficial* No. 49.523.

—. (26 de mayo del 2015b). Decreto 1074 de 2015, Decreto Único Reglamentario del Sector Comercio, Industria y Turismo. *Diario Oficial* No. 49.523. [Artículo 2.2.2.25.1.3].

—. (26 de mayo del 2015c). Decreto 1080 de 2015, Decreto Único Reglamentario del Sector Cultura. *Diario Oficial* No. 49.523. [Artículos 2.8.2.1.1 y ss.; artículo 2.8.5.2.1].

—. (26 de mayo del 2015d). Decreto 1081 de 2015, Decreto Único Reglamentario de la Presidencia de la República. *Diario Oficial* No. 49.523. [Artículo 2.1.1.1.1 y ss.; artículo 31: Anonimización de información mixta]. [Compilación del Decreto 103 de 2015].

—. (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*.

—. (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. (29 de febrero del 2024). Acuerdo 001 de 2024, por el cual se establece el Acuerdo Único de la Función Archivística. Bogotá D.C.: AGN. Disponible en: 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. (5 de septiembre del 2002). Sentencia T-729 de 2002. M.P. Eduardo Montealegre Lynett.

—. (4 de marzo del 2003). Sentencia C-185 de 2003. M.P. Eduardo Montealegre Lynnet. Fundamento jurídico 9.

—. (8 de marzo del 2004). Sentencia T-216 de 2004. M.P. Eduardo Montealegre Lynett. Fundamento jurídico 21.

—. (2005). Sentencia T-718 de 2005.

—. (2008). Sentencia C-1011 de 2008.

—. (2010). Sentencia T-656 de 2010.

—. (6 de octubre del 2011). Sentencia C-748 de 2011. M.P. Jorge Ignacio Pretelt Chaljub. Fundamentos jurídicos 2.4.3.2 y 2.6.3.

—. (6 de diciembre del 2013). Sentencia T-926 de 2013. M.P. Mauricio González Cuervo.

—. (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.

—. (s.f.). Auto 073.

—. (s.f.). Auto 182.

—. (s.f.). Auto 778. Protocolos de acceso para el Sistema Integral de Verdad.

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. AGN.

Departamento Administrativo Nacional de Estadística (DANE). (2024). Guía para la anonimización de datos estructurados. Bogotá D.C.: DANE – Sistema Estadístico Nacional (SEN). Dirección General: Piedad Urdinola Contreras. Subdirección: Andrea Ramírez Pisco. Coordinación técnica: Edna Patricia Rangel Barragán. Marco normativo de referencia: Ley Estadística 2335 de 2023. Documento elaborado a partir de la experiencia conjunta del DANE y el Archivo General de la Nación (AGN). Disponible en: https://www.dane.gov.co/files/sen/registros-administrativos/guia-anonimizacion-datos2024.pdf

Archivo General de la Nación Jorge Palacios Preciado, y 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. AGN – CNMH. Disponible en: 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: Heidy 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. CNMH. Disponible en: https://centrodememoriahistorica.gov.co/descargas/politica-publica-archivo-ddhh.pdf

—. (s.f.). Guía para la Anonimización de Datos e Información No Estructurada: Estándares y Lineamientos Técnicos. CNMH.

Superintendencia de Industria y Comercio (SIC)

Superintendencia de Industria y Comercio. (2018). Concepto No. 7075 de 2018. [Clasificación de imagen fotográfica y video]. SIC.

—. (2014). Concepto No. 14-268304. Clasificación de datos educativos y académicos. 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]. MinTIC.

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.

Artículos científicos y estudios técnicos

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

Sweeney, L. (2002). k-anonymity: A model for protecting privacy. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 10(5), pp. 557–570. Disponible en: https://doi.org/10.1142/S0218488502001648

Carlini, N., Tramèr, F., Wallace, E., Jagielski, M., Herbert-Voss, A., Lee, K., Roberts, A., Brown, T., Song, D., Erlingsson, Ú., Oprea, A., y Raffel, C. (2021). Extracting Training Data from Large Language Models. En Proceedings of the 30th USENIX Security Symposium, pp. 2633–2650. USENIX Association. Disponible en: https://www.usenix.org/conference/usenixsecurity21/presentation/carlini-extracting

Carlini, N., Ippolito, D., Jagielski, M., Lee, K., Tramèr, F., y Zhang, C. (2023). Quantifying Memorization Across Neural Language Models. En Proceedings of the 11th International Conference on Learning Representations (ICLR 2023). Disponible en: https://arxiv.org/abs/2202.07646

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

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

Machanavajjhala, A., Kifer, D., Gehrke, J., y Venkitasubramaniam, M. (2007). l-diversity: Privacy beyond k-anonymity. ACM Transactions on Knowledge Discovery from Data, 1(1), Article 3. Disponibñe en: https://doi.org/10.1145/1217299.1217302

Li, N., Li, T., y Venkatasubramanian, S. (2007). t-Closeness: Privacy beyond k-anonymity and l-diversity. En Proceedings of the IEEE 23rd International Conference on Data Engineering (ICDE 2007), pp. 106–115. Estambul: IEEE. Disponible en: https://doi.org/10.1109/ICDE.2007.367856

Rocher, L., Hendrickx, J., y 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

Todt, J., Hanisch, S., y Strufe, T. (2024). Fantômas: Understanding Face Anonymization Reversibility. Proceedings on Privacy Enhancing Technologies (PoPETs), 2024(4), pp. 24–43. Disponible en: https://arxiv.org/abs/2210.10651

Staab, R., Vero, M., Balunović, M., y Vechev, M. (2024). Beyond Memorization: Violating Privacy Via Inference with Large Language Models. ICLR 2024. Disponible en: 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. Disponible en: https://aclanthology.org/2023.trustnlp-1.23/

Sánchez, D., y Batet, M. (2016). C-sanitized: A privacy model for document redaction and sanitization. Journal of the American Society for Information Science and Technology, 67(1), pp. 148–163. Disponible en; https://doi.org/10.1002/asi.23363

Wiepert, D. y otros. (2024). Reidentification of Participants in Shared Clinical Data Sets:Experimental Study. JMIR AI. Disponible en: 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. Disponible en: https://www.aepd.es

Information Commissioner's Office (ICO). (2012, actualizado 2021). Anonymisation: managing data protection risk – code of practice. Wilmslow: ICO. Disponible en: https://ico.org.uk/media/1061/anonymisation-code.pdf

Agencia Española de Protección de Datos (AEPD). (2016). Orientaciones y garantías en los procedimientos de Anonimización de datos personales. Madrid: AEPD. Disponible en: https://www.aepd.es/guias/guia-orientaciones-procedimientos-anonimizacion.pdf

Agencia Española de Protección de Datos (AEPD), y 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. Disponible en: https://www.gov.br/anpd/pt-br/assuntos/noticias/anpd-abre-consulta-a-sociedade-sobre-o-guia-de-anonimizacao-e-pseudonimizacao

Estándares técnicos internacionales (ISO / NIST)

International Organization for Standardization e International Electrotechnical Commission. (2018). ISO/IEC 20889:2018. Privacy enhancing data de-identification terminology and classification of techniques. Ginebra: ISO. Disponible en: https://www.iso.org/standard/69373.html

National Institute of Standards and Technology. (2015). NIST Internal Report 8053: De-Identification of Personal Information (NIST IR 8053). Gaithersburg, MD: U.S. Department of Commerce, NIST. Disponible en: https://doi.org/10.6028/NIST.IR.8053

National Institute of Standards and Technology. (2014). NIST Special Publication 800-88 Revision 1: Guidelines for Media Sanitization. Gaithersburg, MD: U.S. Department of Commerce, NIST. Disponible en: https://doi.org/10.6028/NIST.SP.800-88r1

National Institute of Standards and Technology. (2023). NIST Special Publication 800-188: De-Identifying Government Datasets. Gaithersburg, MD: U.S. Department of Commerce, NIST. Disponible en: https://doi.org/10.6028/NIST.SP.800-188

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). ONU.

Open Society Justice Initiative. (2013). Principios Globales sobre Seguridad Nacional y el Derecho a la Información (Principios de Tshwane). Open Society Foundations.

ARTICLE 19. (1996). Principios de Johannesburgo sobre Seguridad Nacional, Libertad de Expresión y Acceso a la Información. ARTICLE 19. Disponible en: 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). OCDE. Disponible en: 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. ONU. Disponible en: https://undocs.org/A/RES/68/167

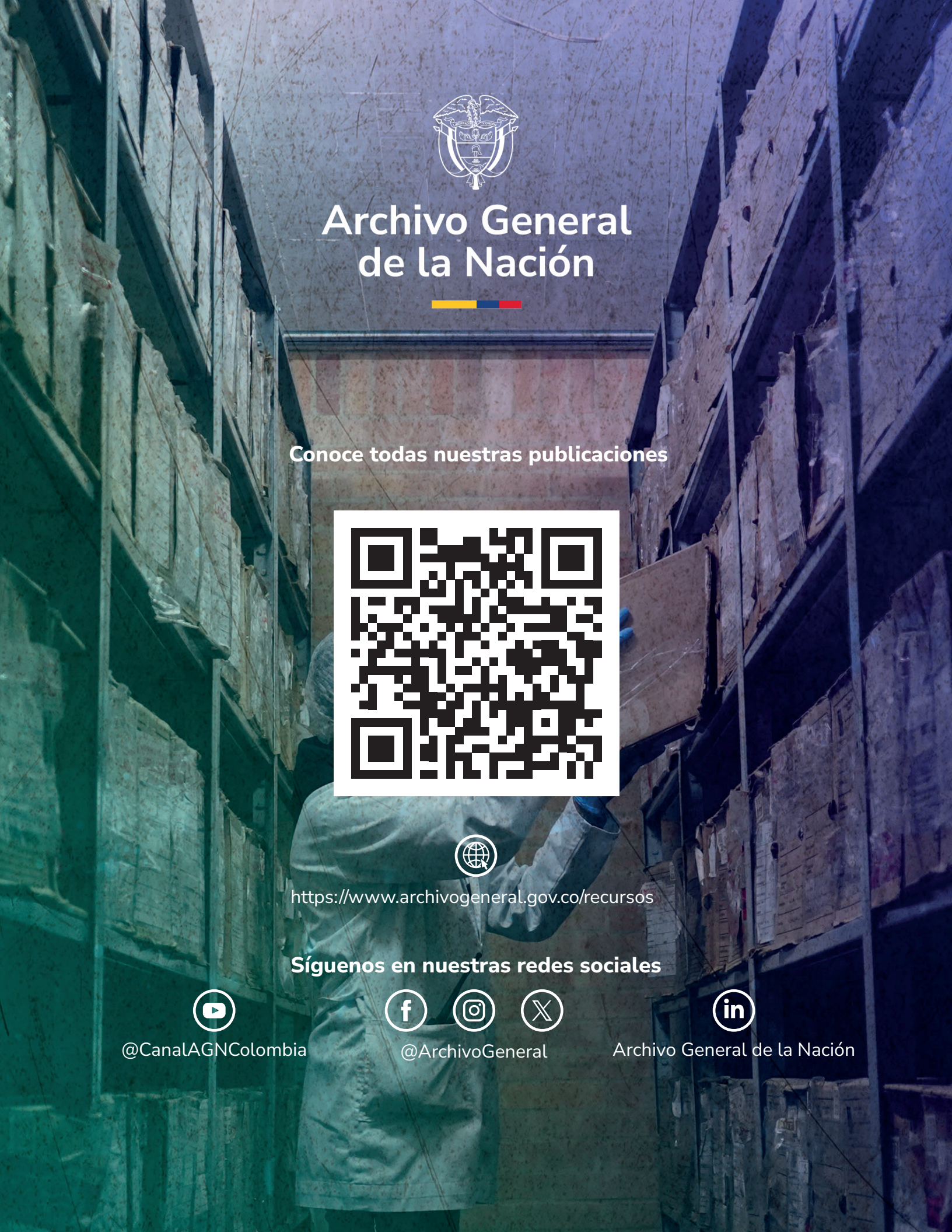
Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos (OHCHR). (2014). El derecho a la privacidad en la era digital. Informe A/HRC/27/37. ONU. Disponible en: https://undocs.org/A/HRC/27/37

Ochoa, S., Rasmussen, J., Robson, C., y Salib, M. (2007). Reidentification of Individuals in Chicago's Homicide Database: A Technical and Legal Study. Massachusetts Institute of Technology (MIT). Disponible en: http://web.mit.edu/sem083/www/assignments/reidentification.html

Data Protection Commission (DPC) de Irlanda. (2019). Guidance on Anonymisation and Pseudonymisation. Dublín: DPC. Disponible en: 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. Disponible en: https://www.adobe.com/es/acrobat/resources/ocr.htmle



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