

Guía metodológica para el fortalecimiento de la gestión automatizada de la información médica mediante la homogeneización y anonimización de datos.

Proyecto



Saludtech Santander

Documento de Trabajo – Versión 2.0

Fecha: 9 de Septiembre de 2025

Elaborado por: Juan Carlos Abaunza, Consultor Líder – AIATIC

Contacto: +57 3232037789 | info@aiatic.com



EJECUTADO POR:



OPERADO POR:



ESTRATEGIA:



aiatic.solutions



+57 3232037789



@aiaticom

Las grandes ideas nacen
de cosas sencillas

INTRODUCCIÓN

La digitalización del sector salud representa una oportunidad sin precedentes para mejorar la eficiencia operativa, la calidad de la atención al paciente y la investigación biomédica. La gestión automatizada de la información médica es el pilar fundamental de esta transformación. Sin embargo, este avance conlleva una responsabilidad inherente: el manejo ético y seguro de datos personales sensibles, que exige un estricto cumplimiento normativo.

La **homogenización** (o estandarización) y la **anonimización** se erigen como procesos técnico-metodológicos cruciales para abordar estos retos. La primera garantiza que los datos sean consistentes, interoperables y de alta calidad. La segunda, mediante técnicas robustas, protege la privacidad de los individuos, permitiendo el uso analítico de la información minimizando el riesgo de re-identificación.

La gestión automatizada de la información médica ofrece grandes ventajas para la eficiencia del sistema de salud, la investigación y la toma de decisiones. Sin embargo, el manejo de datos personales sensibles exige un estricto cumplimiento de la normativa de protección de datos. La homogenización y anonimización son procesos cruciales para garantizar la privacidad y seguridad de la información, permitiendo su uso responsable para fines legítimos.

Esta guía se desarrolla en el marco del **Proyecto Saludtech**, dirigido a las **12 entidades del clúster de salud de Santander**, con el objetivo de superar los desafíos críticos que enfrenta la región y el país:

- **Heterogeneidad tecnológica y fractura digital:** La coexistencia de sistemas de información legacy (como HIS) con soluciones modernas, sumada a brechas de infraestructura y capacidades técnicas, genera silos de información que impiden una visión unificada del paciente y del sistema.
- **Fragmentación de Datos:** La información clínica, administrativa y financiera reside en formatos y estándares dispares, dificultando su integración, análisis y explotación para la inteligencia sanitaria.
- **Marco Normativo Complejo:** El cumplimiento de la Ley 1581 de 2012 (Protección de Datos) y su decreto reglamentario 1074 de 2015 es no solo un mandato legal, sino un



aiatic.solutions



+57 3232037789



@aiaticom

requisito para generar confianza y habilitar el uso secundario de los datos para investigación y salud pública.

1. OBJETIVOS ESTRATÉGICOS

1.1. Objetivo General

Fortalecer la gestión automatizada de la información médica en las 12 entidades beneficiarias del proyecto, mediante la implementación de un marco metodológico robusto de homogenización y anonimización de datos, que garantice su interoperabilidad, calidad, seguridad y uso ético.

1.2. Objetivos Específicos

1. **Estandarizar y unificar:** Establecer y implementar estándares de datos comunes (sintácticos y semánticos) para homogenizar la información proveniente de los diversos sistemas fuentes de las entidades.
 2. **Proteger la privacidad:** Diseñar e implementar un proceso de anonimización basado en riesgo, asegurando el cumplimiento de la normativa de protección de datos personales y minimizando el riesgo de re-identificación.
 3. **Garantizar la utilidad:** Asegurar que los conjuntos de datos homogenizados y anonimizados conserven su valor analítico y utilidad para los fines de investigación, salud pública y mejora de la gestión clínica definidos.
 4. **Fomentar la interoperabilidad:** Crear los cimientos técnicos y operativos para un intercambio seguro y eficiente de información entre las entidades del clúster y con el sistema de salud nacional.
 5. **Crear capacidades sostenibles:** Desarrollar competencias técnicas y de gobernanza al interior de las entidades para la gestión continua y la mejora iterativa de los procesos de datos.
 6. **Cerrar brechas digitales:** Reducir las asimetrías tecnológicas y de conocimiento entre las entidades del clúster mediante capacitación, acompañamiento y la promoción de estándares abiertos.
- **Nota:** La guía es escalable y puede servir como referencia para otras entidades del SNS.



aiatic.solutions



+57 3232037789



@aiaticom

2. ALCANCE

2.1. Entidades Incluidas:

Esta guía se aplica a todas las entidades que manejan información médica en Colombia, específicamente a las 12 empresas del clúster de salud de Santander beneficiarias del proyecto objeto de esta consultoría. Incluyendo:

- Instituciones Prestadoras de Servicios de Salud (IPS) públicas y privadas.
- Entidades Administradoras de Planes de Beneficios (EAPB).
- Entidades territoriales de salud.
- Centros de investigación y universidades.
- Entidades gubernamentales que manejan información de salud.

2.2. Tipos de datos cubiertos:

- **Datos clínicos:** Historia clínica electrónica (HCE), diagnósticos (CIE-10), procedimientos, signos vitales, notas de evolución, resultados de laboratorio (LIS)
- **Datos administrativos:** Datos de identificación, afiliación, autorizaciones, facturación (CUPS).
- **Datos epidemiológicos:** Notificaciones de enfermedades de interés en salud pública.

2.3. Procesos Cubiertos:

El alcance abarca el ciclo completo de vida de los datos, desde su origen hasta su uso analítico:

1. Extracción desde sistemas fuentes.
2. Homogenización, limpieza y estandarización.
3. Almacenamiento seguro (en reposo y en tránsito).
4. Proceso de anonimización/seudonimización.
5. Gobierno, acceso y uso de los datos anonimizados.
6. Auditoría y monitorización continua.



aiatic.solutions



+57 3232037789



@aiaticom

3. MARCO LEGAL EN COLOMBIA

Es fundamental el cumplimiento de las siguientes leyes y regulaciones:

- **Constitución Política de Colombia:** Artículos sobre el derecho a la intimidad y habeas data.
- **Ley Estatutaria 1581 de 2012 (Protección de Datos Personales):** Regula el tratamiento de datos personales, incluyendo los datos sensibles como los de salud.
- **Decreto 1074 de 2015 (Decreto Único Reglamentario del Sector Comercio, Industria y Turismo):** Reglamenta la Ley 1581 de 2012.
- **Ley 1266 de 2008 (Habeas Data Financiero):** Aunque enfocada en datos financieros, establece principios importantes sobre el manejo de información personal.
- **Ley 2015 de 2020** que creó la Historia Clínica Electrónica Interoperable (HCEI) para facilitar el intercambio seguro de datos de salud entre los diferentes prestadores.
- **Código Nacional de Buenas Prácticas para las Estadísticas Oficiales (CNBP) del DANE:** Promueve la calidad y credibilidad de las estadísticas oficiales, incluyendo las de salud.

4. METODOLOGÍA DETALLADA: ENFOQUE POR ETAPAS

4.1. Fase 0: Gobernanza y Conformación del Equipo Interdisciplinario

Un equipo centralizado de gobierno del dato es crítico para el éxito.

- **Roles Clave:**
 - **Sponsor ejecutivo:** Alta dirección de la entidad o del clúster.
 - **Oficial de protección de datos (DPO/Responsable de Tratamiento):** Vela por el cumplimiento legal.
 - **Gestor de datos de salud (Health Data Steward):** Conoce los datos clínicos y sus flujos.
 - **Arquitecto de datos:** Diseña la infraestructura y los modelos.
 - **Ingeniero de datos:** Implementa los pipelines de ETL.
 - **Especialista en ciberseguridad:** Asegura la infraestructura.
 - **Bioestadístico / científico de Datos:** Valida la utilidad de los datos anonimizados.

- **Representantes de usuarios finales (investigadores, epidemiólogos):** Definen los requisitos de uso.

4.2. Fase 1: Identificación y clasificación de activos de datos

- **Inventario y Mapeo:** Catalogar todos los sistemas fuentes, bases de datos, tablas y campos relevantes.
- **Clasificación de Atributos:**
 - **Identificadores Directos:** Nombre, cédula, email, dirección IP. *Deben ser suprimidos o seudonimizados de manera irreversible en la mayoría de los casos.*
 - **Identificadores Indirectos (Cuasi-Identificadores):** Fecha de nacimiento, código postal, sexo, fecha de admisión. *Requieren técnicas de anonimización como generalización o perturbación.*
 - **Datos Sensibles:** Diagnóstico, tratamiento, información genética, resultados de laboratorio. *Nunca deben ser anonimizados por sí solos; siempre deben protegerse en conjunto con los cuasi-identificadores.*
 - **Datos No Sensibles:** Datos anonimizados o agregados, métricas de servicio no asociadas a individuos.

5.3. Fase 2: Homogenización y Estandarización de Datos

- **Definición de un metamodelo de datos común:** Establecer un modelo de datos de referencia para el clúster (por decisión de la fábrica de software basado en HL7 FHIR).
- **Estandarización sintáctica:**
 - Fechas (ISO 8601: AAAA-MM-DD).
 - Booleanos (Sí/No, 1/0).
 - Unidades de medida estandarizadas (ej., mg/dL, cm).
- **Estandarización semántica:**
 - Adopción de terminologías clínicas: **CIE-10** para diagnósticos, **CUPS** para procedimientos, **SNOMED CT** para términos clínicos generales (a futuro), **LOINC** para laboratorios (a futuro)
 - Creación de un diccionario de datos centralizado y accesible.
- **Calidad de Datos:**
 - Implementación de reglas de validación (ej., rangos válidos para creatinina, formato de cédula).

- Procesos de limpieza: deduplicación, corrección de errores tipográficos, imputación de valores faltantes (con documentación).
- **Interoperabilidad Técnica:**
 - Uso de **APIs RESTful** basadas en estándares FHIR.
 - Implementación de un **Bus de Interoperabilidad** o una plataforma **iPaaS** para orquestar el intercambio de datos.

4.4. Fase 3: Anonimización Basada en Riesgo

- **Evaluación del Riesgo de Re-identificación:** Analizar el contexto colombiano (fuentes de datos públicas disponibles, motivos de ataque probables) para cada conjunto de datos.
- **Selección de Técnicas:**
 - **Supresión:** Eliminación permanente de identificadores directos. **Dada la naturaleza del proyecto y su alcance, fue esta la técnica elegida para el producto resultante en esta etapa inicial.**
 - **Generalización:** Agrupar valores en categorías (ej., edad: [20-29], [30-39]; ciudad: "Norte de Santander").
 - **Perturbación (Aleatorización):** Añadir "ruido" estadístico a valores numéricos continuos (ej., $\pm 5\%$ a un resultado de laboratorio).
 - **Seudonimización (Tokenización):** Reemplazo reversible de identificadores con un token. **CRÍTICO:** La tabla de correspondencia (token $\rightarrow$ ID real) debe gestionarse de forma ultra-segura, separada de los datos anonimizados, y con acceso restringidísimo. No es anonimización per se.
 - **K-anonymity, L-diversity, T-closeness:** Técnicas estadísticas avanzadas para garantizar que cada registro en un dataset sea indistinguible de al menos k-1 otros registros respecto a sus cuasi-identificadores.
- **Evaluación de la Utilidad Post-Anonimización:** Verificar mediante análisis estadístico que las distribuciones, correlaciones y patrones de los datos se mantienen válidos para su fin analítico.
- **Documentación (Metadatos de Anonimización):** Registrar exhaustivamente cada técnica aplicada, los parámetros usados y el juicio de riesgo/utilidad.

4.5. Fase 4: Gobierno y gestión de datos anonimizados

- **Política de acceso y uso:** Establecer protocolos claros para solicitar, aprobar y acceder a datasets anonimizados. Implementar acuerdos de uso de datos (DUA). **Esto es un**

ejercicio interno de cada entidad de salud como propietaria de los datasets resultantes, los cuales, en esta etapa del proyecto no se almacenan y/o utilizan en otros procesos en la plataforma construida.

- **Seguridad de la información:**
 - Cifrado de datos en reposo (AES-256) y en tránsito (TLS 1.3).
 - Control de acceso basado en roles (RBAC) y principio de mínimo privilegio.
 - Auditoría de acceso: logs detallados de quién, cuándo y por qué accedió a qué datos.
- **Ciclo de vida de los datos:** Definir políticas de retención y disposición segura tanto para los datos anonimizados como para las tablas de seudonimización (esta segunda técnica a futuro)
- **Capacitación continua:** Programas de sensibilización y entrenamiento para todo el personal que manipule datos, desde el administrativo hasta el investigador.

5. HERRAMIENTAS Y TECNOLOGÍAS (CON ÉNFASIS EN LA INTEROPERABILIDAD):

Categoría	Herramientas/Tecnologías (Ejemplos)	Propósito
Almacenamiento y Procesamiento	PostgreSQL, Azure Blob Storage	Almacenar y procesar grandes volúmenes de datos homogenizados.
Orquestación de Datos (ETL/ELT)	Apache Airflow, Apache Spark	Automatizar los pipelines de homogenización y anonimización.
Interoperabilidad y APIs	HL7 FHIR Servers (Azure FHIR Server, API Gateway)	Facilitar el intercambio estandarizado y seguro de datos entre sistemas.
Gobernanza y calidad	Apache Atlas	Gestionar el catálogo de datos, el linaje y las reglas de calidad.



aiatic.solutions



+57 3232037789



@aiaticom

Lenguajes de Programación	Python (Pandas, PySpark, NumPy)	Scripting personalizado para limpieza, análisis y anonimización.
Ciberseguridad	Herramientas de cifrado, Gestión de Identidades y Accesos (IAM)	Proteger la infraestructura y los datos.

6. ESTRATEGIAS PARA EL CIERRE DE BRECHAS

- **Plan de capacitación gradual:** Cursos modulares online y presenciales dirigidos a distintos perfiles: "Protección de datos para personal clínico", "Anonimización técnica para ingenieros", "Gobernanza para directivos".
- **Comunidad de práctica (CoP):** Crear un foro permanente para las 12 entidades donde compartan desafíos, soluciones y mejores prácticas, facilitado por el consultor.
- **Guías de implementación rápidas (Quickstart Guides):** Documentos simplificados y listas de verificación para que las entidades con menos recursos inicien su implementación.
- **Centro de excelencia virtual:** Un portal web con todos los recursos, estándares, herramientas y foros de discusión para el clúster.
- **Pilotos por nivel de madurez:** Implementar proyectos piloto en entidades con alta, media y baja madurez tecnológica para generar casos de éxito replicables y aprender de los desafíos.

7. MONITOREO, EVALUACIÓN Y MEJORA CONTINUA (MEC)

- **Indicadores Clave de Desempeño (KPIs):**
 - *KPI de Calidad:* Porcentaje de datos que pasan las reglas de validación.
 - *KPI de Interoperabilidad:* Número de transacciones exitosas via API FHIR.
 - *KPI de Seguridad:* Número de incidentes de seguridad relacionados con datos.
 - *KPI de Utilidad:* Número de proyectos de investigación/publicaciones que utilizan los datos anonimizados.
- **Auditorías periódicas:** Evaluaciones trimestrales o semestrales para verificar el cumplimiento de las políticas y la efectividad de las técnicas de anonimización.

- **Revisión y actualización del marco:** La guía debe ser un documento vivo, revisado anualmente para incorporar nuevas regulaciones, técnicas de ataque y tecnologías.

8. CONSIDERACIONES FINALES

- La anonimización es un proceso continuo y debe adaptarse a las nuevas tecnologías y regulaciones.
- Es fundamental mantener la transparencia y la rendición de cuentas en el uso de datos anonimizados.
- Se debe promover la capacitación y la sensibilización sobre la importancia de la protección de datos en el sector salud.
- La transformación digital del sector salud requiere una inversión sostenida en infraestructura, tecnología y capacitación.
- La colaboración entre diferentes actores (entidades de salud, gobierno, academia, sector privado) es fundamental para superar las brechas existentes.
- Es importante adoptar un enfoque gradual e iterativo, priorizando la implementación de soluciones que generen un mayor impacto a corto plazo.
- Se debe monitorear y evaluar continuamente el progreso de la implementación de la estrategia de gestión automatizada de la información médica.

Esta guía proporciona un marco metodológico sólido para fortalecer la gestión automatizada de la información médica en Colombia, específicamente para el clúster de Salud de Santander, garantizando la protección de datos y promoviendo el uso responsable de la información para el beneficio de la salud pública.

Crédito: A Colombia Productiva como sponsor del proyecto y fortalecer la transformación en las regiones, a Cámara de Comercio de Bucaramanga, como articulador y operador de esta iniciativa y fomentar el desarrollo del tejido empresarial en Santander. A las entidades de salud beneficiarias por su compromiso, dedicación y confianza.

A todo el equipo de trabajo de A&A Soluciones – TIC SAS BIC en las actividades de consultoría y de todo el ciclo de vida del desarrollo del producto.



aiatic.solutions



+57 3232037789



@aiaticom