



PLATAFORMA SAAS MULTIEMPRESA · LEY 21.719

Anonimiza tus documentos de forma irreversible

TarjaDoc detecta y elimina datos personales y sensibles de documentos PDF, Word, Excel, PPTX, TXT e imágenes, con IA en capas, trazabilidad verificable y cumplimiento de la **Ley 21.719** de Protección de Datos Personales de Chile.

PDF · Word · Excel · PPTX · TXT

JPG · PNG · TIFF con OCR

Detección por capas con IA

Eliminación irreversible verificada

On-premise: tus datos nunca salen

¿Por qué ahora?

La **Ley 21.719** moderniza la protección de datos personales en Chile y define obligaciones concretas para toda organización que trate información de personas naturales. Entra en vigencia el **1 de diciembre de 2026** y trae sanciones que llegan a **20.000 UTM** por infracción gravísima (art. 35).

CONTEXTO LEGAL

La ley define el **dato personal** como «cualquier información vinculada o referida a una persona natural identificada o identificable» (art. 2 letra f) y exige tratar los datos bajo los principios del **art. 3**: licitud y lealtad, finalidad, proporcionalidad, calidad, responsabilidad, seguridad, transparencia y confidencialidad.

Anonimización — art. 2 letra k)

Procedimiento **irreversible** por el cual el dato no puede vincularse a una persona determinada. Un dato anonimizado **deja de ser dato personal**.

Seudonimización — art. 2 letra l)

Los datos ya no se atribuyen a un titular sin información adicional, que debe mantenerse **separada y protegida**. TarjaDoc la guarda en una bóveda cifrada.

Datos sensibles — art. 2 letra g)

Salud, biometría, origen étnico, afiliación política o sindical, creencias religiosas, vida sexual, datos de menores y situación socioeconómica: las **9 categorías** que TarjaDoc detecta.

Principios — art. 3

Licitud y lealtad, finalidad, proporcionalidad, calidad, responsabilidad, seguridad, transparencia e información y confidencialidad.

«El tratamiento de datos personales con fines distintos a los informados queda prohibido: la finalidad es la brújula de todo tratamiento lícito.»

— Principio de finalidad, art. 3 letra b) de la Ley 21.719

CONSEJO

Los documentos históricos y expedientes públicos concentran gran parte del riesgo: tachar el dato visible no alcanza si el texto queda recuperable en metadatos, capas ocultas o adjuntos. La redacción debe ser profunda, no cosmética.

Qué hace TarjaDoc

Un pipeline transparente, de la carga a la auditoría. La detección propone, la persona decide y la redacción solo ocurre tras la aprobación humana.

1

Sube

Carga PDF, Word, Excel, PPTX, TXT o imágenes. Los escaneados pasan por OCR automático con Tesseract.

2

Detecta

La IA en capas identifica datos personales y sensibles: patrones, NER con spaCy y LLM solo para ambigüedades.

3

Trata

Aplica tachado, anonimización, seudonimización o enmascaramiento según las políticas de tu empresa.

4

Audita

Documento irreversible con cadena de hashes verificable y trazabilidad completa de cada tratamiento.

ficha_paciente.pdf

Tratado

NOMBRE

RUT

CORREO

TELÉFONO

DIRECCIÓN

CUENTA

PATENTE

[CORREO]

+56 9

PERSONA-001

✓ Sin datos personales recuperables

INFORMACIÓN CLAVE

El pipeline automático **nunca redacta por sí solo**: la redacción es irreversible y exige aprobación humana sobre los hallazgos ya revisados. Cada tratamiento queda registrado en la auditoría.

TIP

Los documentos se procesan en **colas por empresa**: el volumen de una organización no bloquea a las demás y cada tenant mantiene su aislamiento.

CONSEJO

Configura una **política de redacción por defecto** por tipo de dato (tachado, anonimización, seudonimización o enmascaramiento) para que cada documento se trate de forma consistente.

6

familias de formato: PDF, Word, Excel, PPTX, TXT e imágenes

9

categorías sensibles del art. 2 letra g) detectadas

≥ 0,95

recall meta en categorías sensibles (estándar de calidad interno)

30

días de retención del original por defecto, luego se purga

Tarja Doc

Brochure comercial · 03 / 08

Detección por capas de IA

Cada capa aporta lo que mejor sabe hacer, y la detección nunca se cae por falta de un detector: si una capa no está disponible, el resto sigue.

1

Determinística

Patrones exactos y validados: **RUT con módulo 11**, patentes chilenas, cuentas y tarjetas con **Luhn**, rol de causa, IP, fecha de nacimiento contextual, email y teléfono, más los patrones propios de tu empresa.

2

Gráfica

Identifica regiones candidatas a **firma y timbre** en documentos escaneados, donde el texto no siempre se extrae como tal.

3

NER spaCy

Reconocimiento de entidades en español con el modelo **es_core_news_md**: nombres de persona, organizaciones y lugares en contexto.

4

Inferencial

Heurísticas de contexto que resuelven casos que los patrones no cubren, elevando la confianza del hallazgo antes de descartarlo.

5

LLM

Solo sobre los fragmentos ambiguos que las capas locales no resolvieron con confianza suficiente, y solo si tu empresa lo habilita. El modelo puede ser local (Ollama/Gemma) o una API externa que tú configures.

Catálogo de entidades detectadas

IDENTIFICADORES DIRECTOS

RUT

Nombre

Email

Teléfono

Dirección

Patente

Cuenta bancaria

Rol de causa

N.º expediente

Fecha nacimiento

IP

Foto de identidad

Firma

9 CATEGORÍAS SENSIBLES · ART. 2 LETRA G)

Salud

Biometría

Origen étnico

Afiliación política o sindical

Creencias religiosas

Vida sexual e identidad de género

Datos de menores (NNA)

Situación socioeconómica

Otros sensibles

 **TIP**

Un hallazgo con confianza bajo el umbral no se descarta en silencio: se marca como **low_confidence** para que el revisor decida. El objetivo es no perder recall, sobre todo en las categorías sensibles.

Cinco tratamientos, una política por empresa

Cada tipo de dato se trata según la acción configurada en la política de tu empresa. Antes y después, sobre el mismo documento ficticio.

Tachado
redact_black

ANTES	DESPUÉS
12.345.678-5	

Anonimización
anonymize

ANTES	DESPUÉS
Juan Pérez González	[NOMBRE]

Seudonimización
pseudonymize

ANTES	DESPUÉS
juan.perez@ejemplo.cl	PERSONA-001

Enmascaramiento parcial
mask_partial

ANTES	DESPUÉS
+56 9 8765 4321	+56 9 4321

Enmascaramiento total
mask_full

ANTES	DESPUÉS
001-2345678-90	

CONSEJO

La **seudonimización es reversible**: el nexa original se guarda cifrado en una bóveda (Fernet) y solo un usuario autorizado puede revelarlo. Úsala cuando necesites reidentificar a una persona por una causa legítima; usa la anonimización cuando no.

GARANTÍA DEL RESULTADO

Antes de publicar, TarjaDoc **sanea el artefacto en profundidad** —metadatos, XMP, adjuntos, anotaciones, formularios AcroForm, marcadores, capas OCG y miniaturas— y luego **verifica** que ningún texto redactado quede recuperable. Si algo se escapó, el artefacto se **destruye** en lugar de entregar un documento con fugas.

Lo que hace la diferencia

Seguridad verificable, trazabilidad inmutable y soberanía de tus datos.

ELIMINACIÓN IRREVERSIBLE

Verificada, no declarada

Si un texto redactado quedara recuperable tras el saneamiento, el artefacto se destruye y no se publica. La redacción profunda incluye metadatos, adjuntos, capas ocultas y miniaturas.

AUDITORÍA

Cadena de hashes SHA-256

Cada evento se encadena al anterior formando un log **inmutable**. El informe de cumplimiento lleva un sello de integridad SHA-256, y la auditoría jamás guarda el texto detectado.

SEUDONIMIZACIÓN

Bóveda cifrada reversible

El nexa original se cifra con Fernet y solo puede revelarse vía **POST /vault/reveal** por personal autorizado. Ideal para estudios clínicos y análisis estadísticos.

ON-PREMISE

Tus datos nunca salen de tu infraestructura

TarjaDoc se despliega como stack Docker en tus servidores. La base de datos, el almacenamiento y la cola viven contigo; la información solo sale hacia el modelo de IA local o la API que tú configures.

INTEGRACIÓN

API pública v1 + webhooks

Sube, procesa y descarga documentos por API con claves **td_...** y scopes *documents:read|write|process*. Los webhooks te avisan cuando un documento cambia de estado.

ESCALA

Colas por empresa y OCR multi-formato

Procesamiento masivo con colas aisladas por tenant y OCR con Tesseract para escaneados. PDF, Word, Excel, PPTX, TXT e imágenes JPG/PNG/TIFF.

«Los datos residen donde el cliente despliega la aplicación, no en una nube operada por el proveedor.»

— Decisión de producto SEC-10: auto-hosting on-premise/Docker

TIP

Configura **original_retention_days**: el original se conserva 30 días por defecto y luego se elimina de forma irreversible, en línea con los principios de finalidad y proporcionalidad del art. 3.

ESTÁNDAR DE CALIDAD

El motor se evalúa contra un corpus de referencia con un recall meta **≥ 0,95 en categorías sensibles**. Es un estándar interno de mejora continua, no una garantía contractual.

Límites configurables, sin sorpresas

Cada empresa tiene un plan con cuotas ajustables a su operación. Comienza con un **plan piloto de 3 meses** y escala cuando lo necesites.

DIMENSIÓN	QUÉ LIMITA	CAMPO DEL PLAN
Usuarios	Máximo de usuarios activos por empresa	<code>max_users</code>
Documentos	Documentos procesables por mes	<code>max_documents_per_month</code>
Almacenamiento	Espacio total en MB para la empresa	<code>max_storage_mb</code>
Tamaño por archivo	Peso máximo de cada documento subido (MB)	<code>max_file_size_mb</code>
Páginas por archivo	Máximo de páginas por documento	<code>max_pages_per_document</code>
Piloto	Período de prueba de 3 meses con bloqueo al vencer	<code>is_trial</code>

 **SIN MONTOS OCULTOS**

Los planes se configuran por **límites de uso** (usuarios, documentos, almacenamiento, tamaño y páginas), no por una grilla de precios fija. Habla con el equipo para dimensionar tu plan.

CONSEJO

El **plan piloto de 3 meses** fija la fecha de vencimiento automáticamente y bloquea el acceso al expirar: perfecto para evaluar el flujo completo antes de comprometerte.

«Multiempresa y multiusuario: perfiles y permisos por empresa, con aislamiento de datos entre tenants y administración completa desde el backoffice.»

— TarjaDoc, plataforma SaaS multiempresa



Protege tus documentos hoy mismo

Anonimiza y tacha datos personales con trazabilidad total, eliminación irreversible verificada y cumplimiento de la Ley 21.719.

[Solicitar una demostración](#)

TarjaDoc · Plataforma SaaS de protección de datos personales
On-premise · Docker · Ley 21.719 de Chile
Multiempresa · Auditoría con cadena de hashes · API v1