

GUÍA PARA LA ANONIMIZACIÓN DE BASES DE DATOS



GOBIERNO DE COLOMBIA



EL DEBER DE PROTEGER (RIESGO)

Garantizar la intimidad y
y confidencialidad.

- Constitución Política
(Art. 15): Derecho a la
intimidad y Habeas
Data.

- Ley 1266/2008 &
1581/2012: Protección
de datos personales y
circulación restringida.

- Ley 79/1993:
Reserva Estadística.

EL DEBER DE COMPARTIR (UTILIDAD)

Maximizar el valor público
de los datos.

- Ley 1712/2014:
Transparencia y acceso
a la información pública
(Datos Abiertos).

- Decreto 2573/2014:
Gobierno en Línea
(reutilización de datos).

- Decreto 1743/2016:
Promoción del uso de
microdatos en el SEN.

LA SOLUCIÓN ELEGANTE

PROCESO DE ANONIMIZACIÓN:
EL ÚNICO MECANISMO TÉCNICO QUE
PERMITE LA MÁXIMA DESAGREGACIÓN
ESTADÍSTICA MANTENIENDO LA
CONFIDENCIALIDAD ABSOLUTA.

UN ESTÁNDAR GLOBAL DE CONFIDENCIALIDAD

REINO UNIDO

Oficina del Comisionado de Información: Desarrollo del Código de Buenas Prácticas para la Anonimización. Enfoque en mitigar riesgos de cruce de bases de datos.

ESTADOS UNIDOS

Oficina de Censos: Liderazgo en técnicas de integración y acceso a microdatos para investigación. Uso de umbrales geográficos, infusión de ruido e intercambio de registros.

HOLANDA & UNECE

Estadísticas de Holanda (CBS): Desarrollo del software interactivo μ -ARGUS (microdatos) y τ -ARGUS (datos tabulares) para el control de divulgación.

COLOMBIA (EL SEN)

DANE & MinSalud: Pioneros regionales. Implementación de lineamientos formales y metodologías de reducción/perturbación respaldadas por el Código Nacional de Buenas Prácticas Estadísticas.

EL ECOSISTEMA CONCEPTUAL DE LA ANONIMIZACIÓN

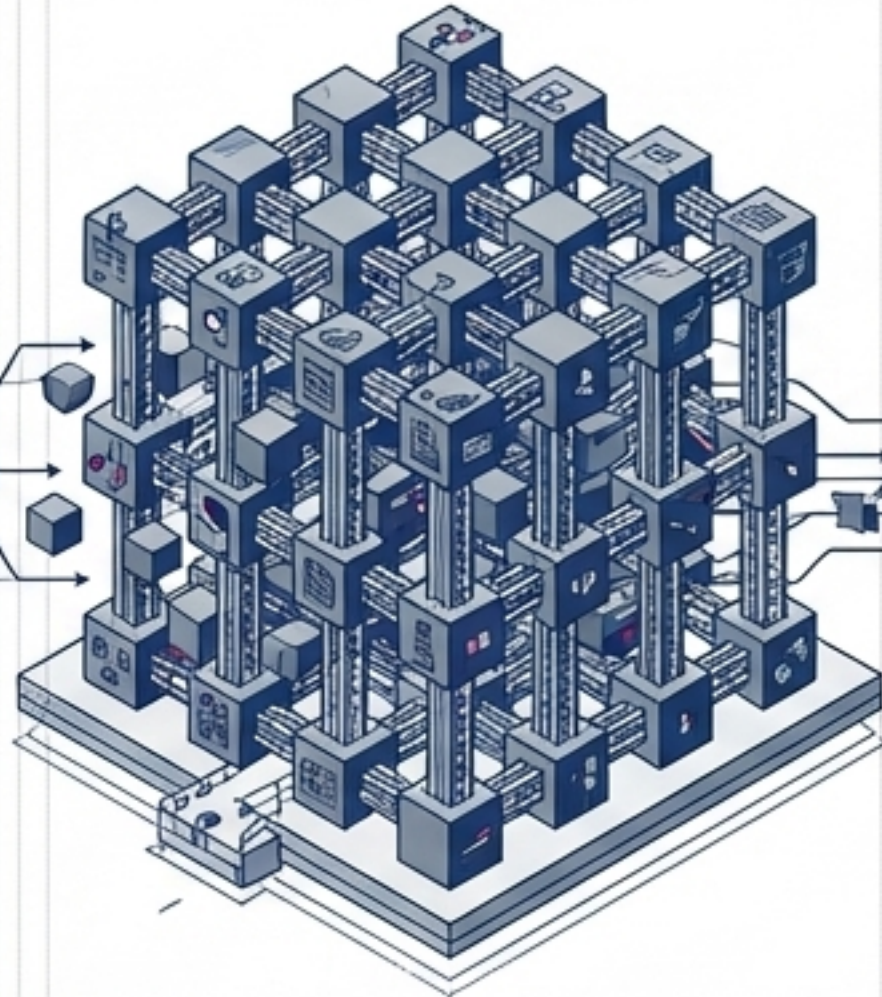
MICRODATO (LA MATERIA PRIMA)

Cada uno de los datos sobre las características de las unidades de estudio (individuos, hogares, establecimientos) consolidados.



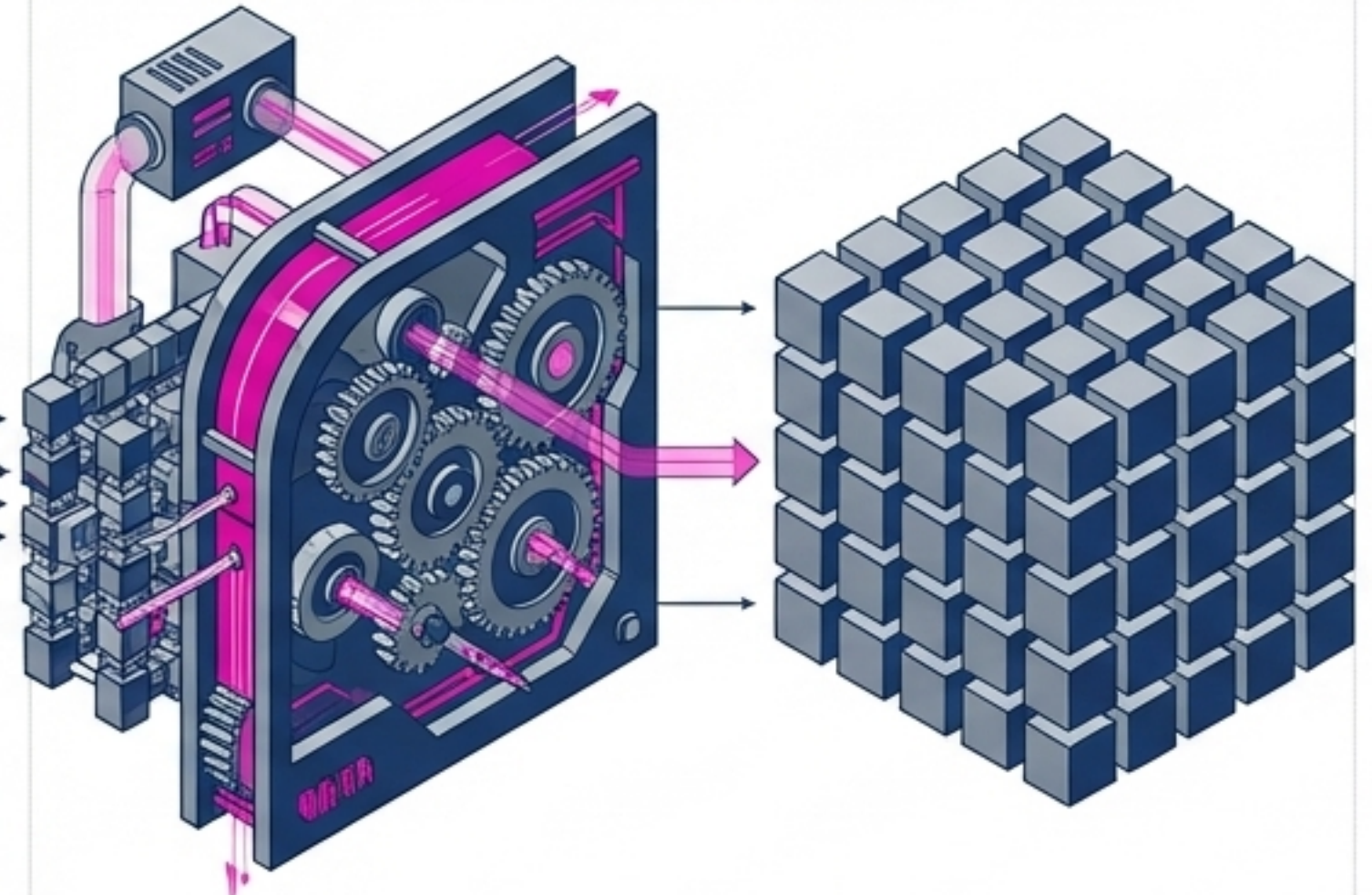
BASE DE DATOS (LA ESTRUCTURA)

Conjunto interrelacionado de datos almacenados sistemáticamente para su posterior uso.

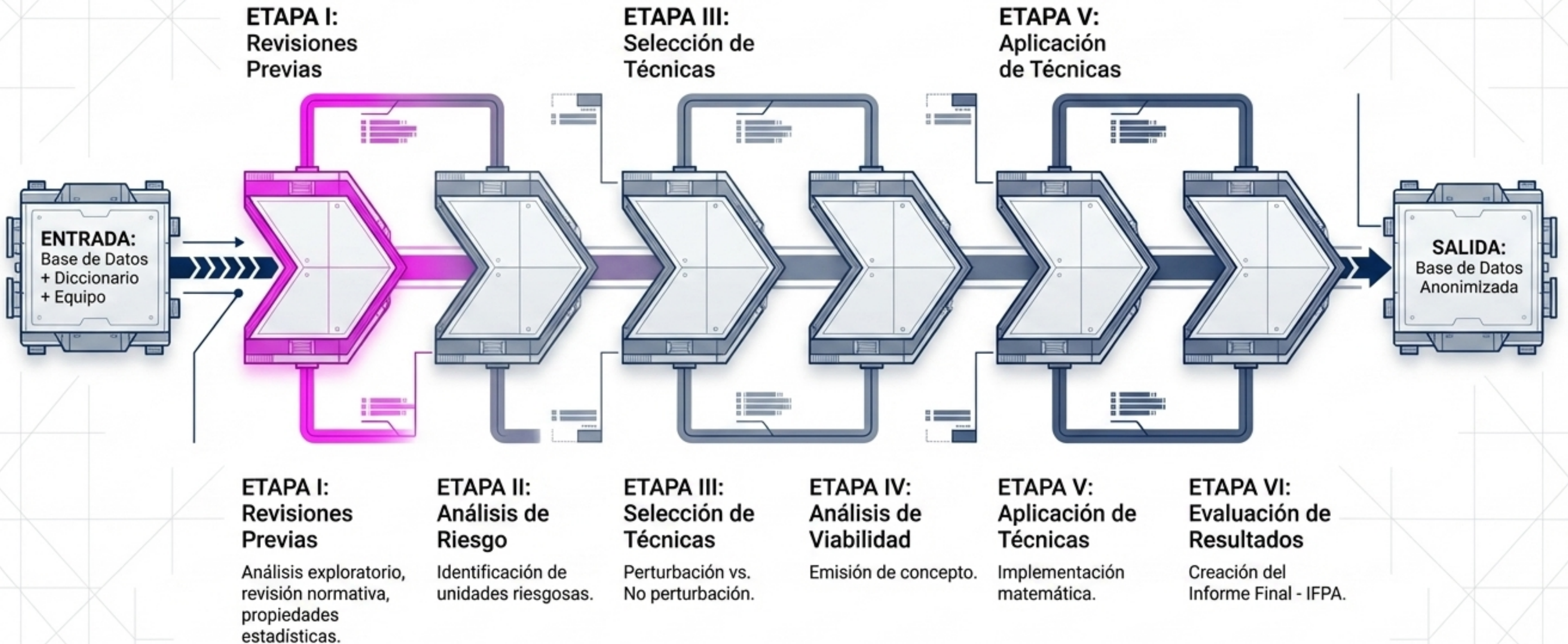


ANONIMIZACIÓN (EL FILTRO DE TRANSFORMACIÓN)

Proceso técnico para transformar datos individuales de tal modo que no sea posible identificar sujetos, preservando las propiedades estadísticas.



EL PLANO ARQUITECTÓNICO: PROCESO DE 6 ETAPAS



TABLERO DE PRE-VUELO: REQUISITOS INICIALES

EQUIPO HÍBRIDO



- ✓ **Conocimiento Temático:** Expertos que entiendan el fenómeno real detrás de los datos.
- ✓ **Capacidad Analítica:** Dominio de software estadístico (R, SAS, SPSS, Stata).

DICCIONARIO DE DATOS



- ✓ **Propiedades exactas:** Nombre, longitud, obligatoriedad, reglas de validación y relaciones de tabla.

INFRAESTRUCTURA TECNOLÓGICA



- ✓ Poder de cómputo alineado al volumen de datos. Verificación de compatibilidad de software para grandes dimensiones.

SEGURIDAD Y PROTOCOLOS



- ✓ Acuerdos de confidencialidad firmados por el equipo.
- ✓ Control estricto de accesos, permisos y contraseñas.

PUNTO DE CONTROL: VERIFICACIÓN DEL ORIGEN

¿Cuál es la naturaleza de la fuente de información?



OPERACIÓN ESTADÍSTICA

- ✓ **Validación de consistencia:** Finalización técnica de la Fase de Ejecución.
- ✓ **Variables calculadas:** Confirmar que las variables derivadas ya están integradas.
- ✓ **Imputación:** Asegurar que los valores faltantes han sido procesados.

REGISTRO ADMINISTRATIVO

- ✓ **Corte temporal estricto:** Definir la versión más actual basada en la periodicidad de consolidación (Ej. Cierre mensual de SISBENNET).
- ✓ **Diagnóstico de calidad:** Ejecutar indicadores de consistencia sobre la captura original.

MATRIZ DE CLASIFICACIÓN DE VARIABLES

		CUANTITATIVAS (CONTINUAS/DISCRETAS)	CATEGÓRICAS
	IDENTIFICACIÓN Datos que apuntan directamente a una unidad (Ej. Cédula, NIT).		
	UBICACIÓN Datos espaciales (Ej. Municipio, Localidad, Departamento).		
	TEMÁTICAS Datos del fenómeno de estudio (Ej. Ingresos, Nivel de escolaridad, Estado civil).		

Toda variable en la base de datos debe ser obligatoriamente clasificada en una de estas intersecciones para determinar su tratamiento estadístico.

DASHBOARD DE PROPIEDADES ESTADÍSTICAS

Métricas base que deben preservarse post-anonimización

Variables Cuantitativas (Propiedades Globales)

	Media	Desviación Estándar	Mínimo	Cuartiles (Q1, Q2, Q3)	Máximo
Ingresos	55.4K	±12.1K	18	[25, 40, 60]	75
Edad	55.4K	±12.1K	18	[25, 40, 60]	75

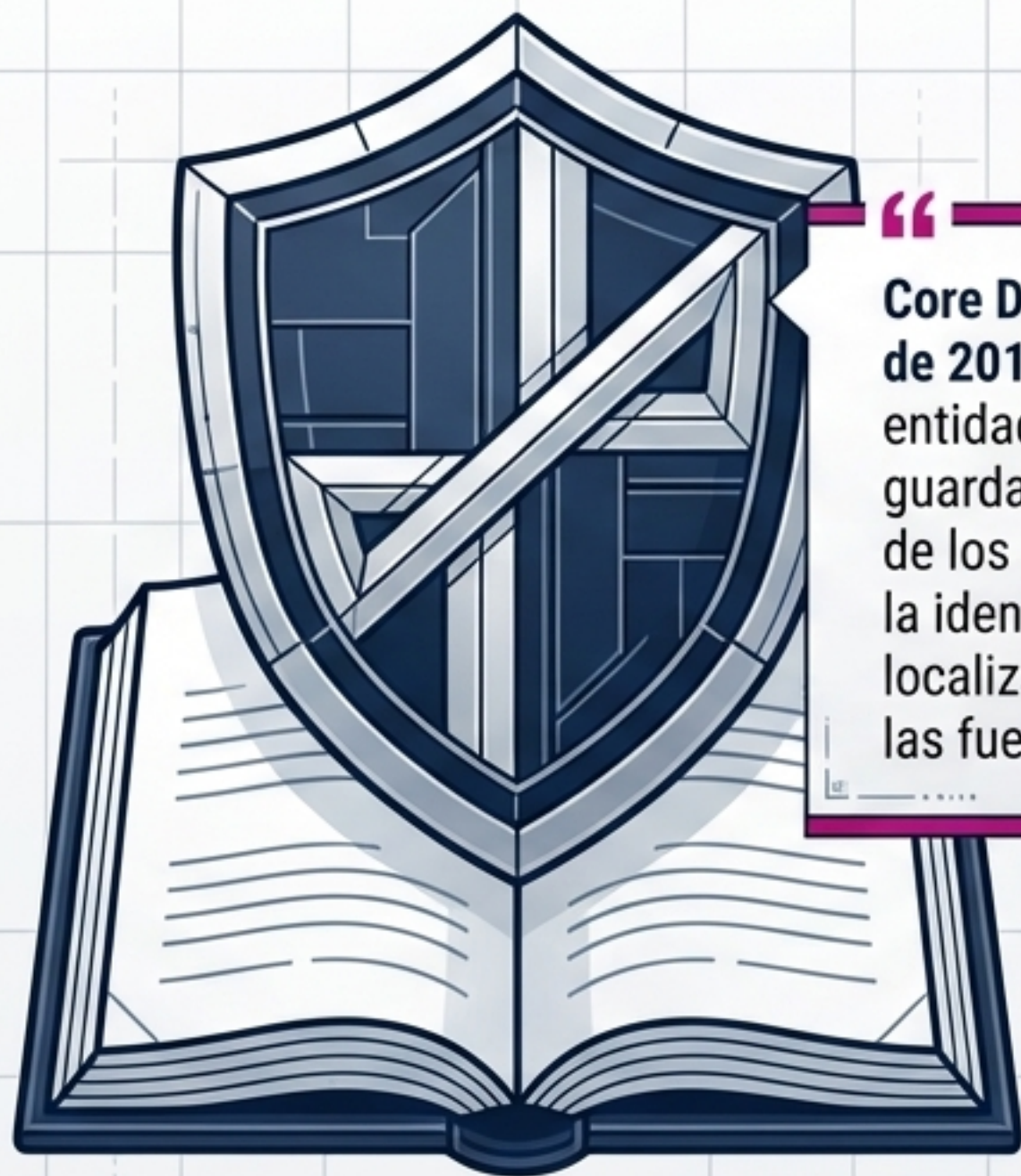
Variables Categóricas (Distribución de Frecuencias)



Estado Civil	Unidades de Observación (N)	Porcentaje (%)
Soltero	1200	60%
Casado	800	40%

Estas métricas son la 'fotografía original'. Servirán como línea base en la Etapa VI (Evaluación) para asegurar que el proceso no destruyó la utilidad estadística.

AUDITORÍA NORMATIVA: DEFINIENDO LAS FRONTERAS



“

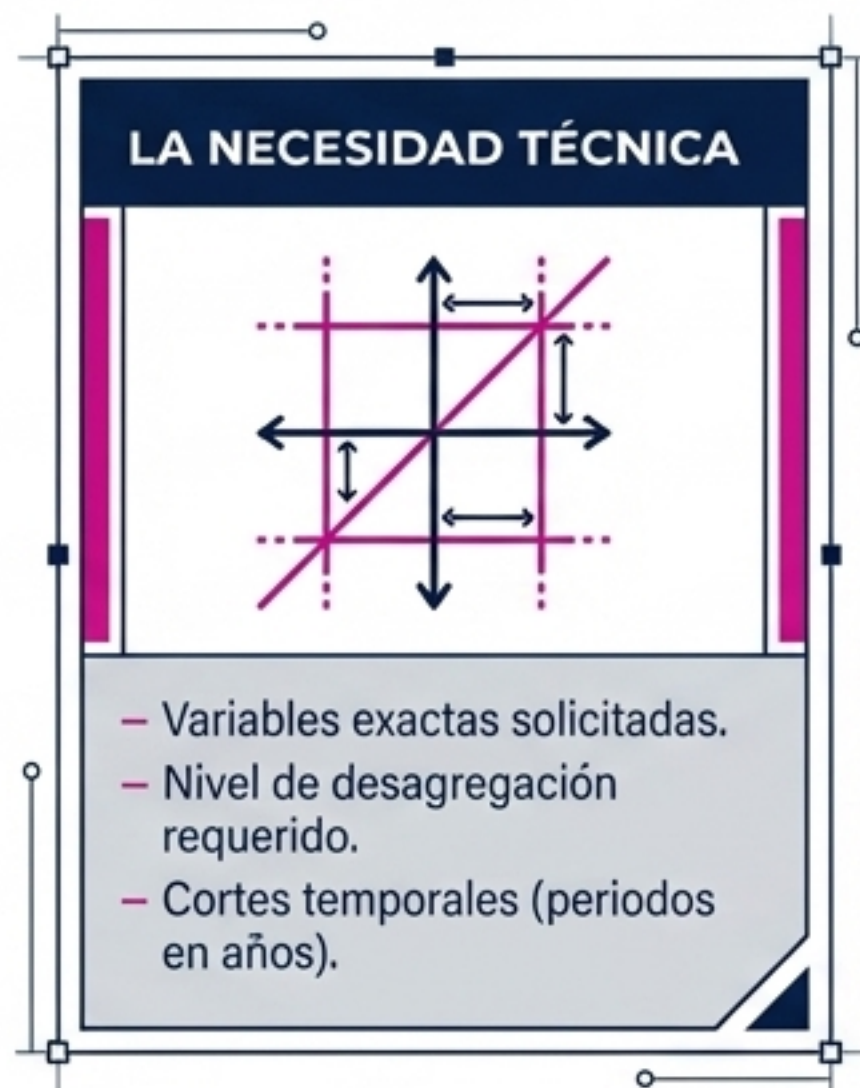
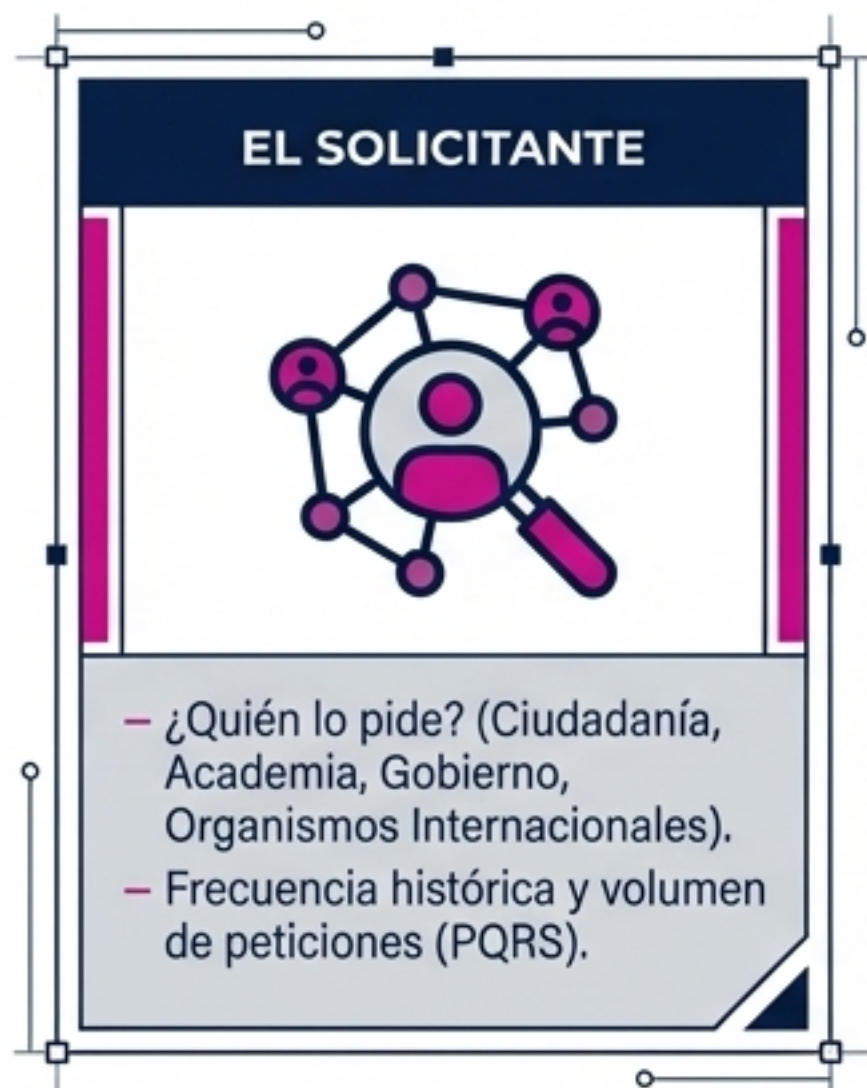
Core Directive (Decreto 1743 de 2016, Art. 2.2.3.3.5): Las entidades del SEN deberán guardar la confidencialidad de los datos que permitan la identificación y/o localización espacial de las fuentes.

LISTA DE VERIFICACIÓN DE AUDITORÍA

- Inventario de Restricciones:** Identificar leyes, decretos y convenios institucionales específicos de la entidad.
- Cláusulas de Reserva:** Verificar estatutos que impidan la publicación directa.
- Respaldos de Publicación:** Documentar normas que exijan apertura de datos (ej. Ley 1712).
- Trazabilidad Final:** Registro formal de quién realizó la revisión legal y en qué fecha para el Informe Final (IFPA).

PERFILAMIENTO DE LA DEMANDA DE INFORMACIÓN

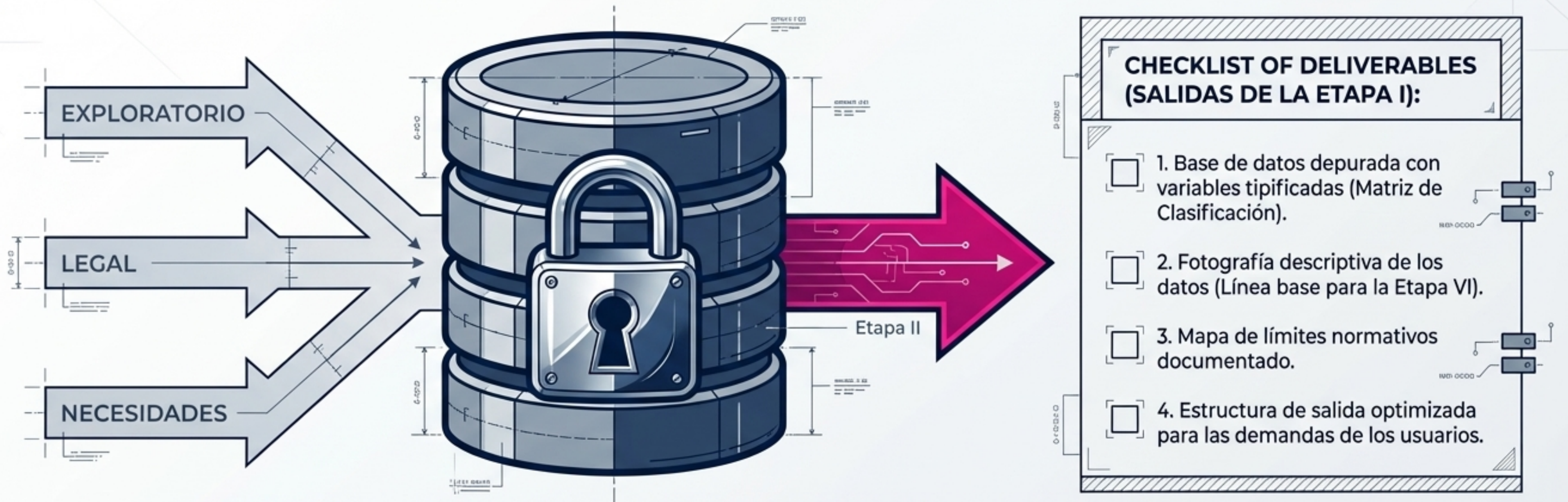
Diseñando para el valor público y reduciendo la carga operativa



Beneficio Estratégico: Publicar bases pre-anonimizadas basadas en este perfilamiento elimina cargas operativas repetitivas y responde de manera proactiva al 80% de los Derechos de Petición.

EL ENTREGABLE DE LA ETAPA I

Preparación completa para el Análisis de Riesgo



Conclusión: Con la arquitectura fundacional construida y el perímetro legal trazado, el equipo está listo para ingresar a la **Etapa II: la detección matemática de las unidades de observación vulnerables a la re-identificación.**