

Censo 2024
Publicación de microdatos

Normativa que rige la protección de información para el INE

SECRETO ESTADÍSTICO · Ley 17.374, artículo 29

«[El INE y sus funcionarios] no podrán divulgar los hechos que se refieren a personas o entidades determinadas de que hayan tomado conocimiento en el desempeño de sus actividades».

Su infracción constituye delito, sancionado por el Código Penal.

Protección de datos personales

Ley 19.628 sobre protección de la vida privada.

Transparencia

Ley 20.285 sobre acceso a la información pública.

Proteger datos hoy es más difícil que en 2017

	Censo 2017	Censo 2024
Preguntas del cuestionario	**21**	**50**
Acceso a inteligencia artificial	Limitado	**Masivo**
Poder de cómputo	Acotado	**Mayor disponibilidad**
Datos externos para cruzar	Menos disponibles	**Abundantes (registros, redes sociales, etc.)**

Más preguntas, más combinaciones.

Cada combinación nueva puede volver única a una persona.

Un censo no es una encuesta.

La cobertura del censo busca ser total y eso un intruso lo sabe.

Datos sin nombre: ¿son anónimos?

Nombre	Edad	Sexo	Comuna	Religión
-	20-24	H	13101	Católica
-	60-64	M	13101	Ninguna
-	35-39	M	13101	Evangélica

Al responder el censo hay personas que entregaron información que no comentan con cualquiera, porque corresponden a su esfera privada.

El INE debe proteger esta información, y al mismo tiempo, publicar parte de ella para que el país la use.

¿Cómo se cumplen ambas promesas?

Respuesta intuitiva: «si no tiene nombre, nadie puede saber quién es».

¿Qué publicó el INE del Censo 2024?

Tres productos que describen a las **mismas personas**.

1

Base de microdatos

Tres tablas relacionadas: viviendas, hogares y personas (~18,5 millones de personas).

2

Base manzana-entidad

189 variables (60 indicadores) agregadas por manzana, entidad rural y otros niveles.

3

Bases cartográficas

Dieciséis bases regionales y una base nacional.

Publicación 4 de diciembre, 2025

Publicación 16 de diciembre, 2025

Fuente: censo2024.ine.gob.cl/resultados

Publicar es equilibrar riesgo y utilidad



Riesgo **Utilidad**

Más detalle da más utilidad, pero también más riesgo.
No existe riesgo cero.
Tampoco es posible la no pérdida de utilidad.

Formas de liberar microdatos: más detalle exige más control de acceso →

Archivo de uso público

Acceso abierto, sin autorización. Menor detalle.

Archivo de uso científico

Acceso con acuerdos y usuarios acreditados.

Enclave de datos

Acceso en entorno controlado. Máximo detalle.

▲ El único que, por normativa, el INE Chile puede publicar

Principios que orientaron el proceso

Utilidad

Maximizar el uso de los datos dentro de lo que permite un archivo público.

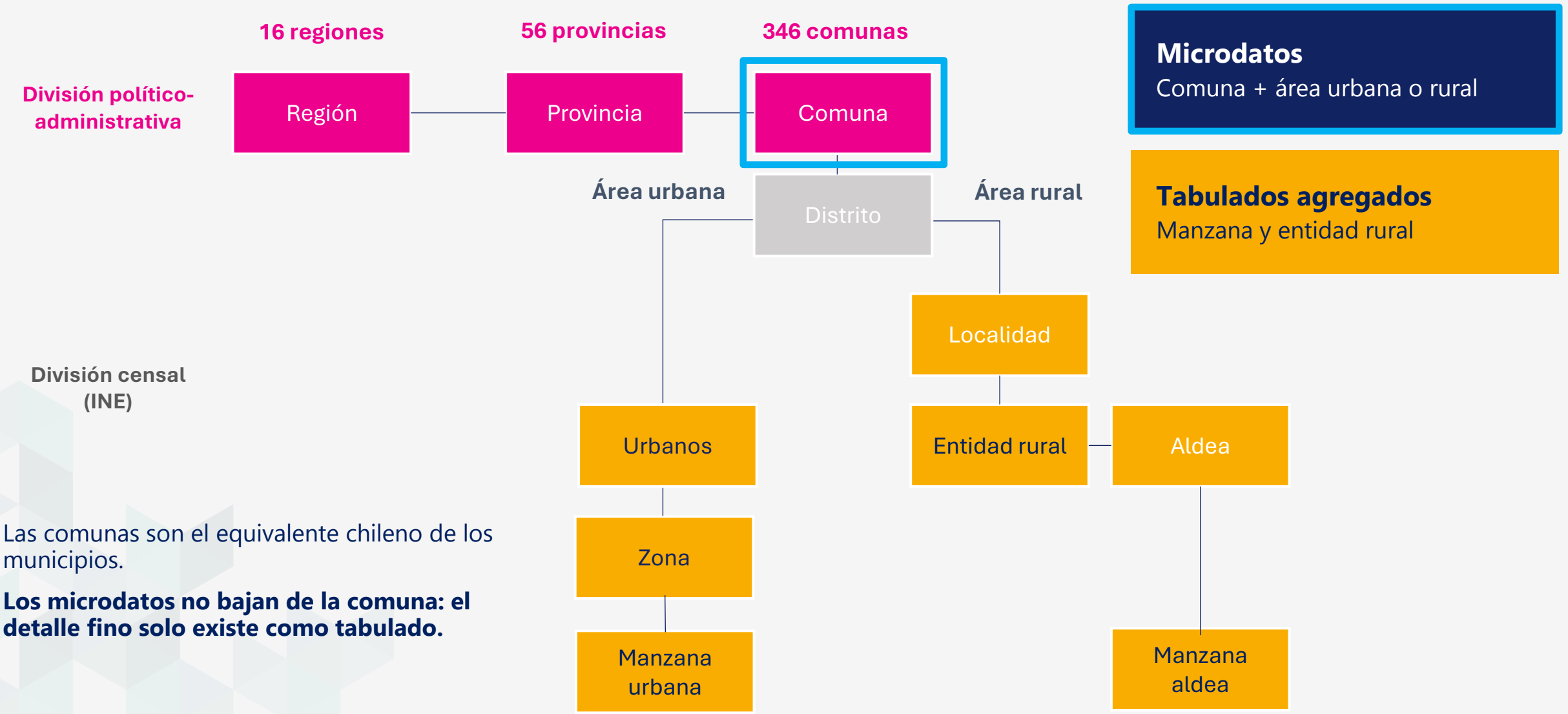
Seguridad

Mantener un nivel adecuado de protección para cada persona.

Simplicidad

Métodos fáciles de explicar, automatizables y eficientes.

¿Hasta dónde bajar en el territorio?



Las comunas son el equivalente chileno de los municipios.

Los microdatos no bajan de la comuna: el detalle fino solo existe como tabulado.

Proceso de control de divulgación de microdatos

Siguiendo la Guía de control de divulgación de microdatos del INE (2021), basada en lineamientos Banco Mundial y en la herramienta sdcMicro.

1. Clasificar las variables
2. Definir escenarios de divulgación
3. Medir el riesgo
4. Aplicar métodos de protección
5. Evaluar los resultados

Microdatos: k-anonimato, medido con sdcMicro (freqCalc) / Tabulados: frecuencia mínima, concentración y supresiones secundarias

Premisa: el riesgo nunca es cero. El objetivo es reducirlo a un nivel aceptable.

Clasificación de las variables

Identificadores directos

Dicen quién eres.

Por ejemplo: Nombres, direcciones, coordenadas geográficas.

Se eliminan

Identificadores indirectos

Solos no te identifican; combinados, sí.

Por ejemplo: Sexo, edad, comuna, nivel educativo, nacionalidad.

Se agrupan u ocultan

Variables sensibles

Lo que alguien podría descubrir de ti.

Por ejemplo: pueblos originarios, creencias religiosas, género.

Protección reforzada

¿Qué dice la ley? La Ley 19.628 y la Ley 21.719, consideran sensibles los datos que revelan el origen étnico o racial, las creencias religiosas y la identidad de género.

Cuando un registro es único, es vulnerable

Ej. Una persona es vulnerable cuando su combinación de **edad, sexo y lugar de residencia** es única.

Datos censales sin tratamiento

Edad	Sexo	Lugar de residencia	Religión
21	H	13101011001001	Católica
24	H	13101011001001	Ninguna
22	H	13101011001001	Evangélica o protestante
22	M	13101011001001	Testigo de Jehová
24	M	13101011001001	Evangélica o protestante
20	M	13101011001001	Otra

Otra base de datos disponible (ficticia)

Edad	Sexo	Lugar de residencia	Nombre	RUN
21	H	13101011001001	Juan Pérez	22.123.456-7
20	M	13101011001001	Catalina González	21.555.555-5

← **Misma edad, sexo y lugar de residencia: es la misma persona.**

Al unir ambas bases por edad, sexo y manzana, se descubre que Juan Pérez es católico.

Cada combinación aparece una sola vez: **k = 1**. Eso es un registro único.

Agrupar para esconderse en la multitud

Si publicamos **tramos de edad** y **comuna** en lugar de edad exacta y manzana-entidad, cada combinación se repite.

Datos censales agrupados

Edad	Sexo	Lugar de residencia	Religión
20-24	H	13101	Católica
20-24	H	13101	Ninguna
20-24	H	13101	Evangélica o protestante
20-24	M	13101	Testigo de Jehová
20-24	M	13101	Evangélica o protestante
20-24	M	13101	Otra

¿Cuál de los tres hombres de 20 a 24 años es Juan Pérez?

Ya no se puede saber con certeza: **k = 3**.

Meta para Censo 2024: k = 5

Recomendación del Banco Mundial. Este ejemplo, con k = 3, todavía no cumple.

k-anonimato: cada combinación de identificadores indirectos debe repetirse al menos k veces en los datos.

Cada agrupación protege más, pero también borra detalle.

Qué se pierde en utilidad para proteger los microdatos

Evaluamos el riesgo en cinco escenarios de divulgación:

Transversal | **Educación** | **Migración internacional** | **Vivienda** | **Trabajo**

Algunos ejemplos de aplicación de métodos no perturbativos:

Edad

- Se trunca en **85 años y más**.
- Comunas de 6.000 habitantes o más (295 comunas): **edad simple**.
- Comunas más pequeñas: **solo tramos de cinco años**.

País de nacionalidad

- Seis países con categoría propia: **Argentina, Perú, Bolivia, Colombia, Venezuela y Haití**.
- El resto de los países se agrupa **por continente**.
- Comunas más pequeñas: **solo Chile o extranjero**.

Otras variables tratadas

- Lugar de nacimiento: tratamiento similar a país de nacionalidad.
- Pertenencia a pueblos originarios.
- Religión o credo.
- Género.
- Etc.

Costo: hay análisis que antes eran posibles para los usuarios y ahora no, sobre todo en las 51 comunas con menos de 6.000 habitantes.

Ejemplos de solicitudes tras la publicación

> **Población por edad simple y sexo en zonas urbanas y localidades rurales.**

> **Población por sexo y edad simple, de 0 a 100 años, para cada manzana y entidad rural.**

> **Hogares unipersonales de personas mayores a nivel de manzana.**

Cada solicitud, por sí sola, parece razonable. Pero cada nueva suma a lo ya publicado.

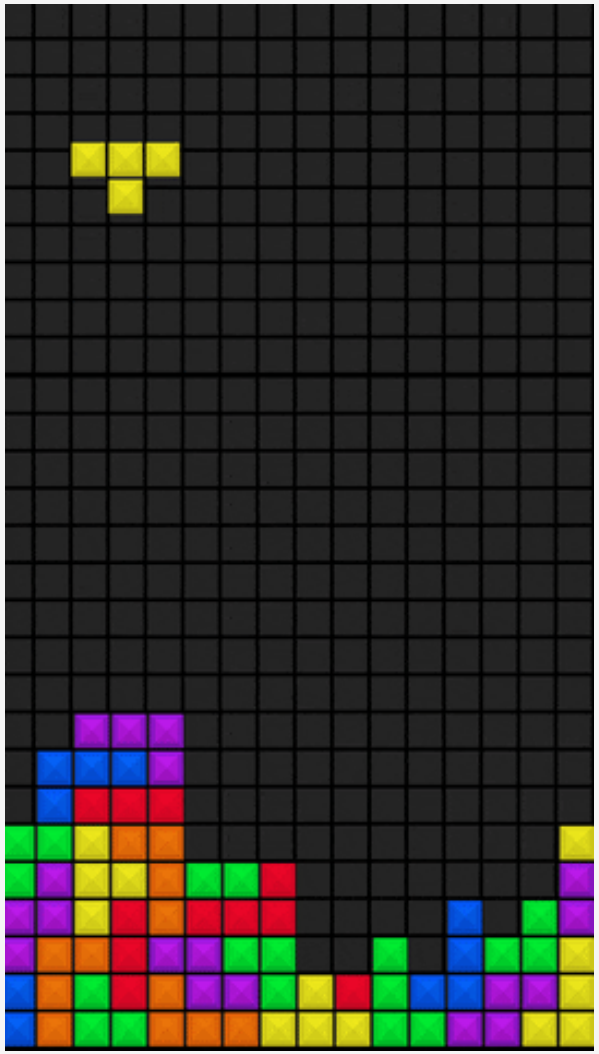
Cada nueva solicitud es evaluada con cautela.

El riesgo se acumula: cada tabla es una pista

Cada tabla, variable o cruce nuevo agrega una pista sobre las mismas personas.

Ej. Un tabulado por manzana funciona como una **ecuación**: fija cuántas personas con ciertas características viven en cada lugar. Con suficientes ecuaciones, se pueden despejar las incógnitas.

Por eso, microdatos y tabulados deben analizarse de manera conjunta al evaluar nuevas publicaciones.



Base de microdatos de personas, hogares y viviendas

Comuna-área

×

Tabulados agregados con 189 variables

Manzana-entidad / Zona-localidad

- Regla de frecuencia mínima
- Regla de concentración
- Supresiones secundarias

EE. UU.: microdatos reconstruidos desde tablas

En el censo de 2010, EE. UU. publicó más de **150 mil millones de estadísticas**, la mayoría por manzana (*census block*). Un equipo de la propia Oficina del Censo intentó reconstruir los registros individuales usando solo esas tablas.

34 de 180

conjuntos de tablas publicados bastaron para el ataque.

70 %

de las manzanas con todos sus registros reconstruidos de forma exacta: 97 millones de personas.

3,4 millones

de personas con raza y etnia inferidas con 95 % de acierto.

Conclusión: una tabla agregada no protege por el solo hecho de estar agregada.

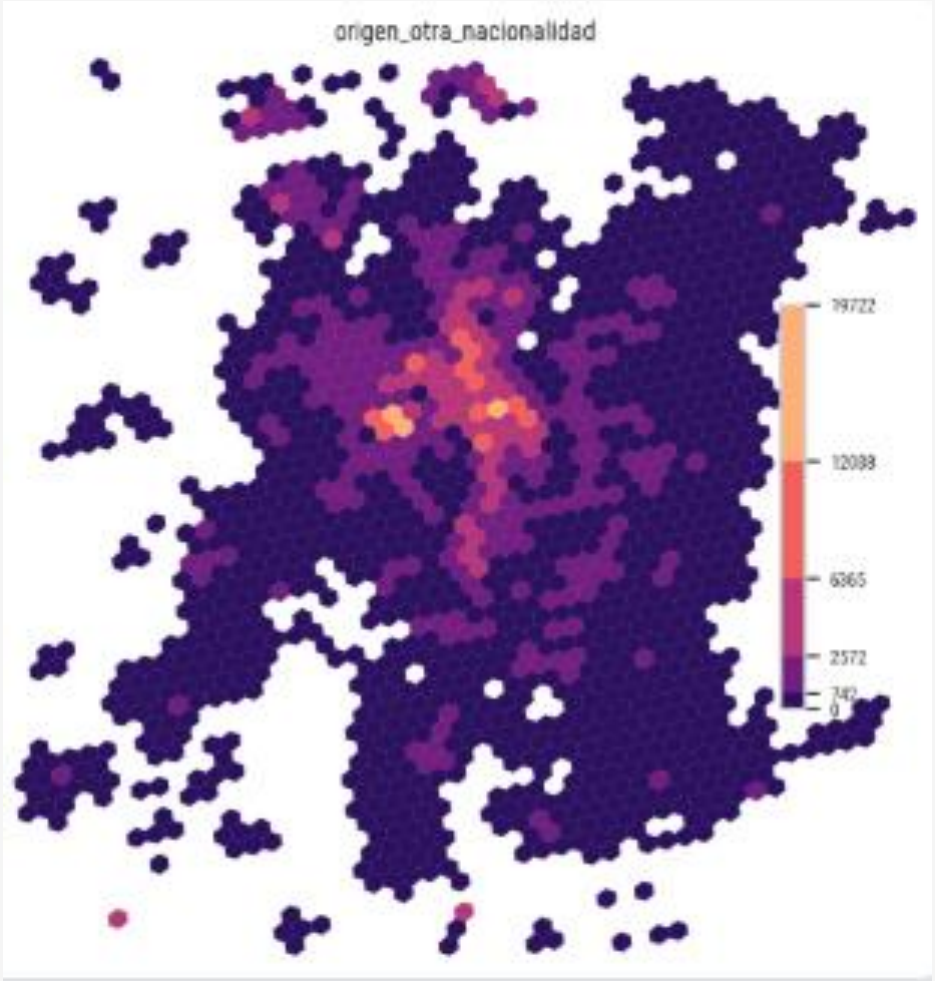
Para el censo de 2020, EE. UU. adoptó la privacidad diferencial.

Fuente: Abowd y otros (2025). A Simulated Reconstruction and Reidentification Attack on the 2010 U.S. Census. CES Working Paper 25-57.

¿Qué ha ocurrido en Chile?

A comienzos de 2026, el académico Eduardo Graells-Garrido publicó un método para ubicar, de manera probabilística, las viviendas de los microdatos dentro de cada comuna.

- Un trabajo motivado por una necesidad legítima de análisis urbano.
- **No es una reidentificación,** pero muestra el camino que se podría recorrer para asignar, de manera probabilística, microdatos a desagregaciones más finas.



Fuente: Graells-Garrido, E. (2026). «Cuando los datos no tienen ubicación: un método para asignar viviendas censales». datagramas.cl

Lo que aprendimos y lo que viene

Tres aprendizajes

1. **El riesgo es acumulativo:** cada nueva publicación debe evaluarse considerando lo ya publicado.
2. **El problema no es teórico:** la experiencia internacional y nacional lo confirma.
3. **El entorno cambia rápido:** los métodos que usamos se adecuan al estándar actual; ante un entorno cambiante, ya debemos explorar los siguientes pasos.

Dime qué dicen los datos y no podré decirte quién eres

Cómo publicar información sensible sin comprometer la privacidad de las personas

Toro, M. «Dime qué dicen los datos y no podré decirte quién eres». Revista Bits de Ciencia, DCC, U. de Chile.

Lo que viene

- **2027:** hoja de ruta de control de divulgación para el próximo censo.
- **Explorar** métodos perturbativos para microdatos y tabulados, incluida la privacidad diferencial.
- **Tender puentes con las universidades,** que ya abrieron esta conversación.

La confidencialidad no es un obstáculo para el acceso: es la condición para que las personas nos sigan abriendo la puerta.

¡Muchas gracias!

INE Instituto Nacional de Estadísticas Chile

CENSO 2024

Anexo: Guía de control de divulgación (INE, 2021)

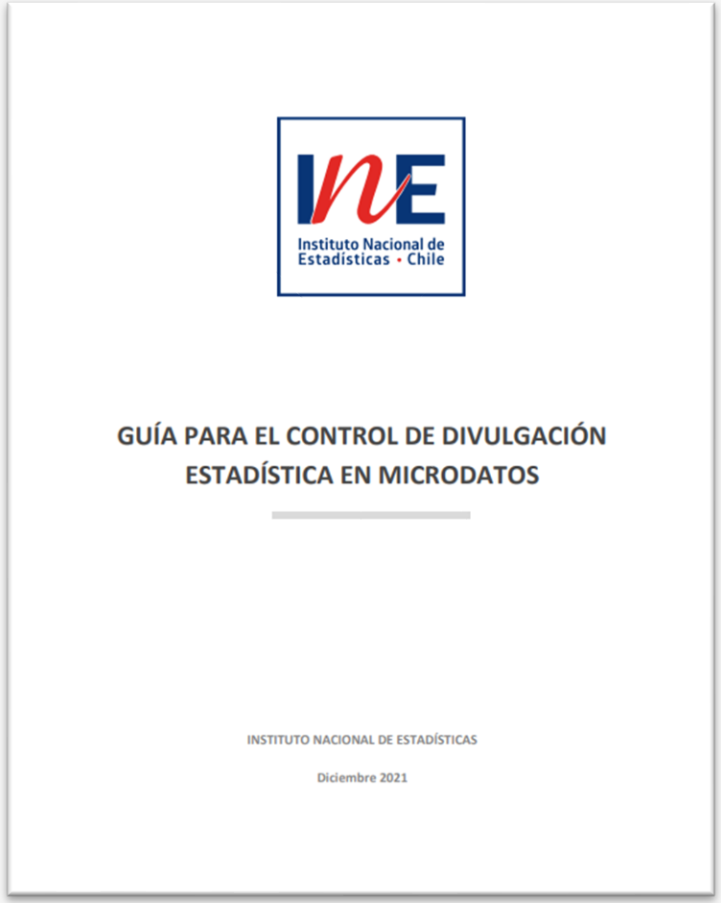
Publicada por el INE en diciembre de 2021.

La guía se basó en el trabajo de:

- Benschop, Machingauta y Welch (2021). Statistical Disclosure Control: A Practice Guide.
- Templ, Kowarik y Meindl (2015). Statistical Disclosure Control for Micro-Data Using the R Package *sdcMicro*. Journal of Statistical Software, 67(4), 1-36. doi:10.18637/jss.v067.i04.

Actualmente el INE se encuentra trabajando en un **nuevo documento**, que incluirá además el control de divulgación en tabulados y datos geoespaciales.

La publicación de los microdatos censales se apoyó en esta guía para estructurar el proceso de control de divulgación.



Fuente: ine.gob.cl, sección Buenas prácticas y estándares.

Anexo: Secreto estadístico y marco legal

Secreto Estadístico

Ley Orgánica 17.374 de creación del Instituto Nacional de Estadísticas

Artículo 29° - El Instituto Nacional de Estadísticas, los organismos fiscales, semifiscales y Empresas del Estado, y cada uno de sus respectivos funcionarios, **no podrán divulgar los hechos que se refieren a personas o entidades determinadas de que hayan tomado conocimiento en el desempeño de sus actividades**.

El estricto mantenimiento de estas reservas constituye el "Secreto Estadístico". Su infracción hará incurrir en el delito pre visto y penado por el artículo 247, del Código Penal, debiendo en todo caso aplicarse pena corporal.

Artículo 30° - Los **datos estadísticos no podrán ser publicados o difundidos con referencia expresa a las personas o entidades a quienes directa o indirectamente se refieran**, si mediare prohibición del o los afectados.

Privacidad y Transparencia

Ley 19.628: protección de la vida privada y el tratamiento de los datos personales en Chile.

Ley 20.285: Ley de Transparencia.