

Instructivo de anonimización y uso de inteligencia artificial

Encuesta a usuarios de bibliotecas públicas 2025





INSTRUCTIVO DE ANONIMIZACIÓN
SISTEMA NACIONAL DE BIBLIOTECAS PÚBLICAS
VERSIÓN INSTITUCIONAL DEFINITIVA
Agosto 2026

Ministerio de las Culturas, las Artes y el Patrimonio

Subsecretaría del Patrimonio Cultural

Servicio Nacional del Patrimonio Cultural

Sistema Nacional de Bibliotecas Públicas

Índice

1. Objeto y alcance	3
2. Principios rectores	3
3. Marco normativo y de referencia	3
3.1 Régimen de protección de datos personales	3
3.2 Uso de inteligencia artificial en la Administración del Estado	4
3.3 Orientaciones complementarias	4
4. Fuente y corte temporal	4
5. Control de consentimiento	4
6. Clasificación de variables y tratamiento aplicado	5
7. Variables críticas que conservan su semántica	5
8. Productos generados	5
9. Control de riesgo de reidentificación	6
10. Muestra analítica	7
11. Flujo de trabajo con inteligencia artificial	7
12. Estado del procedimiento	7
13. Referencias normativas y técnicas	8

1. Objeto y alcance

Este protocolo documenta, de manera reproducible, las decisiones adoptadas para proteger los datos personales de la Encuesta a usuarios de bibliotecas públicas del Sistema Nacional de Bibliotecas Públicas (SNBP) 2025, construir las bases de análisis y establecer las condiciones de uso de herramientas de inteligencia artificial generativa. El procedimiento se encuentra ejecutado y cerrado en todas sus etapas.

El documento distingue entre el archivo fuente, la base pseudonimizada de uso restringido, los productos agregados de difusión y la base anonimizada por casos que fue sometida a control de riesgo. Esa distinción es deliberada: la eliminación de identificadores directos no equivale por sí sola a anonimización, por lo que ninguna base se denomina «anónima» mientras subsista riesgo residual de identificación.

El Sistema Nacional de Bibliotecas Públicas entrega este protocolo a disposición de las demás unidades del Servicio Nacional del Patrimonio Cultural como precedente de trabajo aplicable a otros procesos de levantamiento y publicación de datos.

2. Principios rectores

- **Finalidad.** Los datos se utilizaron exclusivamente para el diagnóstico y el análisis institucional de la encuesta.
- **Anonimización por casos.** Cada base conserva únicamente la información necesaria para la finalidad analítica que le corresponde.
- **Separación de ambientes.** Los datos identificables y pseudonimizados permanecen en entorno local e institucional; los productos destinados a inteligencia artificial superaron controles adicionales.
- **Trazabilidad.** El archivo fuente no fue modificado y cada transformación se ejecutó mediante scripts versionados.
- **No exposición innecesaria.** Nombres, correos electrónicos, direcciones IP, teléfonos, identificadores y textos libres no se compartieron con herramientas de inteligencia artificial generativa.
- **Prudencia.** Ante riesgo residual de identificación, se optó siempre por el producto agregado antes que por el microdato.

3. Marco normativo y de referencia

3.1 Régimen de protección de datos personales

El tratamiento de datos personales se sujetó a la Ley N° 19.628, sobre protección de la vida privada, en su versión vigente hasta el 30 de noviembre de 2026. La Ley N° 21.719, publicada en 2024, establece un nuevo régimen de protección de datos personales y crea la Agencia de Protección de Datos Personales, con entrada en vigencia el 1 de diciembre de 2026. El procedimiento adoptó desde su diseño criterios de anonimización, finalidad, seguridad y responsabilidad compatibles con el estándar reforzado que entrará en vigor.

3.2 Uso de inteligencia artificial en la Administración del Estado

El Oficio Circular N° 711 del Ministerio de Ciencia, Tecnología, Conocimiento e Innovación establece lineamientos para el uso de herramientas de inteligencia artificial en el Estado y recomienda no introducir información personal, especialmente sensible, en herramientas de inteligencia artificial generativa que no hayan sido contratadas por el órgano ni desarrolladas por o para éste. En consecuencia, el procedimiento privilegió el análisis local y la transferencia a inteligencia artificial únicamente de resultados agregados o de productos previamente sometidos a control de riesgo.

3.3 Orientaciones complementarias

Se consideraron las recomendaciones del Consejo para la Transparencia relativas al tratamiento de datos personales por los órganos de la Administración y a la transparencia en el uso de sistemas algorítmicos.

4. Fuente y corte temporal

Elemento	Resultado
Fuente	Archivo SPSS (.sav) exportado desde SurveyMonkey
Período definido	Del 22 de septiembre al 22 de octubre de 2025, 23:59:59
Registros originales	4.165
Variables originales	234
Fecha mínima observada	22 de septiembre de 2025, 09:02:59
Fecha máxima observada	22 de octubre de 2025, 21:21:41
Registros fuera del corte	0

El archivo fuente se conserva sin modificaciones. La verificación estructural y temporal se ejecutó localmente mediante Python, sin exponer respuestas individuales.

5. Control de consentimiento

El control de la pregunta de consentimiento arrojó 3.995 casos con respuesta afirmativa y 170 con respuesta negativa. No se registraron casos sin clasificación.

Las bases analíticas derivadas se construyeron únicamente con quienes aceptaron participar. Los 170 casos sin consentimiento permanecen exclusivamente en el archivo fuente resguardado y no se incorporaron a ningún producto analítico ni de difusión.

6. Clasificación de variables y tratamiento aplicado

Grupo	Ejemplos	Tratamiento aplicado
Identificadores directos	Identificador de respuesta, identificador de recopilador, dirección IP, correo, nombre, apellido y contacto para sorteo	Eliminados de todas las bases derivadas
Textos abiertos	Alternativas «otro» y comentario final	Excluidos íntegramente, sin inspección de contenido individual
Fechas exactas	Fecha de creación y de modificación	Conservadas solo para controles locales
Geografía detallada	Comuna	Excluida de los microdatos; la región se informa únicamente agregada

Grupo	Ejemplos	Tratamiento aplicado
Edad	Edad exacta	Conservada solo localmente; transformada en tramos para los productos de difusión
Datos sensibles	Género, salud y discapacidad, pertenencia indígena	Analizados localmente y difundidos solo en forma agregada con supresión de celdas
Otros cuasi-identificadores	Ocupación, nacionalidad, sexo	Reservados para análisis local y difundidos en forma agregada
Variables sustantivas	Uso, evaluación, lectura, colecciones y servicio principal	Conservadas por ser necesarias para el diagnóstico

7. Variables críticas que conservan su semántica

- **Confianza institucional.** El código 0 corresponde a «No conoce / No aplica». Se conserva como categoría especial y no se confunde con una evaluación de 1 a 7.
- **Evaluación de atributos del servicio.** El código 8 corresponde a «No aplica». Se mantiene separado de los valores 1 a 7 y de los valores perdidos del sistema.
- **Servicio de mayor asistencia.** Se conserva porque identifica el servicio bibliotecario al que la persona declara haber asistido con mayor frecuencia durante el último año. De esta variable se derivó la clasificación analítica de cuatro grupos: biblioteca regional, biblioteca en convenio, Bibliometro y bibliomóvil.
- **Contribución percibida de la biblioteca.** Se corrigió su lectura: no corresponde a una escala numérica, sino a las categorías «Aumentar», «Mantener», «Disminuir» y «No ha influido».

8. Productos generados

- **Base local pseudonimizada restringida.** Base de uso local e institucional. Elimina identificadores directos y textos abiertos y conserva las variables necesarias para el análisis completo. No se comparte con herramientas de inteligencia artificial externas.
- **Base minimizada de datos sometida a control de riesgo.** Versión reducida elaborada para la evaluación de reidentificación. No fue aprobada para difusión ni para uso con inteligencia artificial.
- **Agregados demográficos y sensibles.** Resultados agregados de las variables retiradas de los microdatos, con reglas de supresión de celdas pequeñas.
- **Diccionario de decisiones de minimización de datos.** Registro, variable por variable, de la decisión aplicada y su fundamento.
- **Informe técnico de minimización de datos.** Reporte de ejecución, conteos y evaluación de riesgo.

- **Clave de pseudonimización.** Clave local separada, de custodia restringida, que no se comparte por ningún canal.

9. Control de riesgo de reidentificación

La base minimizada candidata fue sometida a un control de singularidad mediante combinaciones de cuasi-identificadores. El resultado determinó que no era apta para difundirse como microdato ni para utilizarse con herramientas de inteligencia artificial generativa.

Indicador	Resultado
Cuasi-identificadores evaluados	Región, tramo de edad, nivel educativo, tipo de servicio y origen agrupado del recopilador
k mínimo	1
Combinaciones con k menor que 5	794
Registros en combinaciones con k menor que 5	1.289
Combinaciones únicas	500

Un segundo ensayo, con el plan de generalización más restrictivo evaluado —edad agrupada, educación agrupada y tipo de biblioteca—, mantuvo un k mínimo igual a 1, con 10 combinaciones bajo el umbral, 25 registros en ellas y una combinación única. Sobre esa base se adoptó la decisión institucional definitiva de no difundir microdatos individuales y de construir, en su lugar, una base pública agregada de resultados.

4.165 registros en el archivo fuente resguardado	3.822 casos en la base analítica adulta	0 microdatos individuales difundidos
---	---	--

10. Muestra analítica

Entre los 3.995 casos con consentimiento afirmativo se identificaron 3.822 personas de 18 años o más, 22 menores de edad y 151 casos sin edad informada. Un control específico confirmó que los 151 casos sin edad no registraban ninguna respuesta sustantiva posterior, por lo que fueron documentados como abandonos iniciales.

La base analítica principal quedó fijada en 3.822 personas adultas. Los casos de menores se conservan en el archivo fuente y quedan fuera de la muestra analítica, por corresponder este a personas de 18 años o más. Ningún caso fue eliminado con el propósito de reproducir un tamaño muestral utilizado en análisis anteriores.

11. Flujo de trabajo con inteligencia artificial

El procedimiento establecido y aplicado es el siguiente, y rige para futuros procesos del Sistema Nacional de Bibliotecas Públicas:

1. El archivo fuente se conserva en carpeta de acceso restringido y no se modifica.
2. Los scripts de control, minimización de datos y análisis se ejecutan localmente.
3. La base local pseudonimizada y la clave de pseudonimización no se comparten.
4. Los microdatos que no superan el control de riesgo no se difunden ni se transfieren a herramientas externas.
5. A las herramientas de inteligencia artificial se entregan únicamente tablas agregadas, resultados estadísticos, diccionarios y metadatos sin datos personales.
6. Cuando se requiere un análisis nuevo, la herramienta colabora en el diseño del código, que se ejecuta localmente, y solo se devuelve la salida agregada.
7. Se aplica supresión primaria a las celdas de 1 a 4 casos y supresión complementaria cuando un total permita deducir el valor oculto.
8. Cada decisión queda registrada en el diccionario, con control de versiones.

12. Estado del procedimiento

Etapas	Estado
Descarga y corte temporal	Cerrada
Verificación estructural	Cerrada
Control de consentimiento	Cerrada
Eliminación de identificadores directos	Cerrada
Separación de base local y base minimizada	Cerrada
Control de riesgo de reidentificación	Cerrada, con resultado no apto para microdatos
Definición de la base analítica	Cerrada, con 3.822 casos
Construcción de la base pública agregada	Cerrada
Publicación de resultados descriptivos	Cerrada

El alcance de los productos publicados es descriptivo. No se aplicaron factores de expansión ni ponderadores, y las comparaciones entre modalidades de servicio se presentan como diferencias descriptivas.

13. Referencias normativas y técnicas

- Ley N° 19.628, sobre protección de la vida privada, versión vigente hasta el 30 de noviembre de 2026. Biblioteca del Congreso Nacional de Chile.

- Ley N° 21.719, que regula la protección y el tratamiento de los datos personales y crea la Agencia de Protección de Datos Personales. Entrada en vigencia: 1 de diciembre de 2026.
- Ministerio de Ciencia, Tecnología, Conocimiento e Innovación. Oficio Circular N° 711, sobre lineamientos para el uso de herramientas de inteligencia artificial en el Estado.
- Consejo para la Transparencia. Recomendaciones y guías sobre protección de datos personales para los órganos de la Administración del Estado.
- Informe técnico de anonimización y minimización, Encuesta a usuarios de bibliotecas públicas SNBP 2025.

Este documento es un protocolo metodológico y de resguardo de información. No sustituye el pronunciamiento de las unidades jurídica, de seguridad de la información y de protección de datos del Servicio, que deben ser consultadas ante cualquier transferencia de microdatos a terceros.