

Flujo metodológico de anonimización y generación de bases

Encuesta a usuarios de bibliotecas públicas 2025





FLUJO METODOLÓGICO DE ANONIMIZACIÓN Y BASES
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 del documento	3
2. Flujo metodológico ejecutado	3
3. Productos generados y reglas de uso	5
4. Base habilitada para el análisis secundario	5
5. Alcance de la base agregada	6
6. Estado de cierre	6

1. Objeto del documento

Este documento registra, en orden cronológico, las decisiones y los procedimientos aplicados desde la extracción de la base original hasta la construcción de las bases de análisis, los productos de validación interna y los archivos agregados destinados al análisis secundario y al trabajo con herramientas de inteligencia artificial.

Su propósito es asegurar trazabilidad, reproducibilidad, resguardo de los datos personales y claridad respecto de qué archivo corresponde utilizar para cada finalidad. El flujo se encuentra ejecutado y cerrado.



2. Flujo metodológico ejecutado

- 1. Extracción de datos.** Se aplicó un filtro temporal correspondiente al período comprendido entre el 22 de septiembre y el 22 de octubre de 2025, hasta las 23:59 horas. El sistema informó 4.165 respuestas para dicho período.
- 2. Conservación de la fuente original.** La base se descargó en formato SPSS (.sav) y se almacenó sin modificaciones como archivo fuente.
- 3. Control estructural.** Mediante un script ejecutado localmente en Python se verificó que el archivo tuviera 4.165 registros y 234 variables. El procedimiento no modificó la base ni examinó respuestas individuales.

4. Auditoría previa a la minimización de datos. Se ejecutó una auditoría de estructura y riesgo que identificó variables potencialmente identificadoras, respuestas abiertas que contenían datos personales, variables sensibles, códigos especiales de respuesta y mecanismos de recopilación.

5. Control del consentimiento. Dentro del período definido se registraron 3.995 respuestas con consentimiento afirmativo y 170 con consentimiento negativo. Los casos sin consentimiento quedaron excluidos de todas las bases derivadas y se conservan únicamente en el archivo fuente.

6. Primera minimización de datos y desidentificación. Se construyeron bases derivadas eliminando identificadores directos, datos de contacto, textos abiertos y variables técnicas de página. El archivo original no fue sobrescrito. Se crearon variables derivadas para las pruebas de riesgo.

7. Primera evaluación de riesgo de reidentificación. La primera base candidata fue sometida a un control de singularidad. El resultado mostró un k mínimo igual a 1, con 794 combinaciones bajo el umbral, 1.289 registros en ellas y 500 combinaciones únicas. Se determinó que no era apta para difundirse ni para utilizarse con herramientas de inteligencia artificial generativa.

8. Auditoría de completitud. Se examinó el avance por módulos de las 3.995 personas con consentimiento, lo que permitió distinguir entre registros con consentimiento y respuestas efectivamente utilizables.

9. Confirmación de abandonos iniciales y edad. Se identificaron 3.822 personas de 18 años o más, 22 menores de 18 años y 151 casos sin edad. Se confirmó que estos últimos no registraban respuestas sustantivas posteriores y se documentaron como abandonos iniciales.

10. Definición de la base analítica principal. Se fijó como base analítica principal a las 3.822 personas adultas con consentimiento afirmativo dentro del corte temporal. Se estableció que cada pregunta utiliza su propio n válido, sin forzar un denominador común.

11. Generación de la base analítica restringida. Se creó una base de 3.822 casos y 208 variables destinada al análisis estadístico, clasificada expresamente como restringida: no se publica ni se transfiere a servicios externos.

12. Tratamiento de la muestra. La definición de la muestra y selección de la misma fue realizada por el área de estudios, quedando fuera del flujo aquí presente. Los productos aquí descritos son de carácter descriptivo y no incorporan factores de expansión.

13. Generación de resultados descriptivos. A partir de la base analítica adulta se construyó un libro de resultados agregados con n válido, frecuencias y porcentajes por pregunta. El procesamiento cubrió 37 preguntas o grupos, 743 filas de respuestas simples, 156 filas de preguntas múltiples y 31 ítems de escalas o matrices.

14. Aplicación de reglas específicas de codificación. Se documentaron las reglas particulares de las preguntas de confianza institucional y de evaluación de atributos, y se corrigió el tratamiento de la pregunta de contribución percibida, que corresponde a categorías de impacto y no a una escala numérica.

15. Definición de la clasificación analítica por tipo de biblioteca. Se utilizó el servicio principal declarado, con 3.557 respuestas válidas. Se establecieron tres grupos de comparación: bibliotecas públicas regionales, bibliotecas públicas en convenio —que incorporan a los bibliomóviles— y Bibliometro.

16. Generación del libro de resultados por tipo de biblioteca. Se produjo un archivo agregado con el n válido y los resultados del total y de cada modalidad. Las variables con 1 a 4 casos, fueron suprimidas con el fin de respetar la confidencialidad y anonimato de las personas participantes.

17. Generación de la base de validación interna. Se creó una base reemplazando los datos identificadores del participante mediante un código, destinada a que el área de estudios recalculara y corroborara los resultados. Por contener microdatos individuales, circula únicamente por canales institucionales autorizados.

18. Evaluación de una base de datos candidata de microdatos para difusión. Se ensayaron distintos niveles de generalización de cuasi-identificadores. Ninguno alcanzó el umbral de resguardo exigido, por lo que la base de datos candidata fue rechazada de manera definitiva.

19. Decisión de difundir mediante base agregada. En lugar de microdatos individuales se construyó una base agregada cuya unidad de registro es un resultado estadístico y no una persona. Esta estructura reduce de manera sustantiva el riesgo de identificación y habilita el análisis secundario.

20. Elaboración del informe de resultados. A partir del libro agregado se elaboró el informe de resultados, que presenta los resultados generales, los resultados por modalidad de servicio y las orientaciones temáticas para el contraste con la Encuesta Nacional de Participación Cultural y Comportamiento Lector.

21. Generación de la base de revisión interna. Se generó un archivo de revisión con una fila por persona y una columna por variable, con etiquetas legibles y sin identificadores directos ni textos abiertos, destinado exclusivamente a la corroboración técnica interna.

22. Regla operativa sobre uso de inteligencia artificial. Se estableció que los cálculos sobre microdatos se ejecutan localmente en Python y que a las herramientas de inteligencia artificial se entregan únicamente resultados agregados o productos aprobados para difusión.

3. Productos generados y reglas de uso

Producto	Contenido	Uso	Difusión
Archivo fuente	4.165 registros y 234 variables	Conservación sin modificación	No
Base analítica adulta restringida	3.822 casos y 208 variables	Análisis local	No
Base de validación interna	Microdatos desidentificados	Corroboración técnica por	No

Producto	Contenido	Uso	Difusión
		canal institucional	
Base de revisión interna	Una fila por persona, sin identificadores	Revisión del área de estudios	No
Candidata de microdatos	Versión reducida sometida a control de riesgo	Rechazada por riesgo de identificación	No
Libro de resultados por tipo de biblioteca	Resultados agregados con n y porcentajes	Validación y elaboración del informe	Sí
Base pública agregada de resultados	Una fila por resultado estadístico	Análisis secundario y datos abiertos	Sí
Informe de resultados	Lectura narrativa de los resultados	Difusión institucional	Sí

4. Base habilitada para el análisis secundario

El archivo habilitado para solicitar cruces, comparaciones, búsqueda de patrones descriptivos o contextualización con otras fuentes es la base pública agregada de resultados. Cada fila representa un resultado agregado e incluye la pregunta, la modalidad de servicio, la categoría o indicador, el n válido y el porcentaje o media correspondiente. Puede complementarse con el informe de resultados y con los documentos oficiales de las fuentes que se utilicen para el contraste.

No se transfieren a herramientas externas el archivo fuente, la base analítica restringida, la base de validación interna ni la candidata de microdatos que no superó el control de riesgo.

5. Alcance de la base agregada

La base pública agregada permite recuperar los porcentajes de las preguntas de respuesta única y múltiple, la distribución por tipo de servicio y los resultados totales y por modalidad. Para las escalas de 1 a 7 informa la media, el porcentaje de notas 6 y 7 y las categorías especiales, junto con la distribución completa de cada categoría de respuesta.

Los cruces no contenidos en la base agregada se calculan localmente sobre la base restringida y se incorporan como tabla agregada. Este criterio permite ampliar el análisis sin transferir microdatos.

6. Estado de cierre

Hito	Estado
Extracción y resguardo de la fuente original	Cerrado
Control estructural y auditoría previa	Cerrado
Definición de la base analítica adulta	Cerrado
Base analítica restringida	Cerrado
Resultados descriptivos por pregunta	Cerrado
Resultados por tipo de biblioteca	Cerrado
Reglas de codificación de escalas y contribución percibida	Cerrado
Base de validación interna	Cerrado
Base pública agregada	Cerrado
Informe de resultados	Cerrado
Base de revisión interna	Cerrado

Documento institucional del Sistema Nacional de Bibliotecas Públicas. La difusión de microdatos individuales requiere una evaluación institucional específica de riesgo de reidentificación.