



# Nota Metodológica para el Tratamiento de la Base de Nacimientos del Registro Civil

## I. ANTECEDENTES

Como parte de un proceso de transparencia proactiva y mediante los trabajos interinstitucionales de la Agencia Digital de Innovación Pública y el Registro Civil de la Ciudad de México, se acordó la apertura de la base de datos de nacimientos registrados desde el 01 de enero del 2000 hasta la fecha de última actualización al momento de la elaboración de este documento (30 de septiembre de 2022).

La propuesta de apertura en el Portal de Datos Abiertos de la Ciudad de México implica que la base de datos del Registro Civil pasa por un procesamiento que cumple con los siguientes criterios:

- I. Protección de datos personales
- II. Asegurar la fiabilidad de la información
- III. Mantener un estándar de calidad y limpieza de los datos

Con el fin de cumplir con estos tres criterios, la base de datos del Registro Civil fue procesada por el equipo de Datos Abiertos de la Ciudad de México siguiendo los pasos que se describen a continuación.

## II. PROCESAMIENTO DE LOS DATOS

La base de datos de nacimientos registrados en la Ciudad de México puede incluir discrepancias con respecto a los datos publicados por el Instituto Nacional de Estadística y Geografía (INEGI) por algunos motivos que incluyen, pero no son exclusivamente los siguientes:

- Errores en la captura de la información
- Errores en el proceso de digitalización de actas
- Correcciones y cambios en actas de nacimiento

El tratamiento de la información aquí presentado busca solucionar muchos de estos errores, sobre todo en los relacionados con los registros duplicados, sin embargo, para referencias precisas de la natalidad en la Ciudad de México, es recomendable dirigirse a la información generada por el Instituto Nacional de Estadística y Geografía (INEGI).



## Tratamiento de registros duplicados

Siguiendo los lineamientos establecidos por el equipo encargado del proyecto del sistema de actas digitales [www.actas.cdmx.gob.mx](http://www.actas.cdmx.gob.mx) se implementó el siguiente proceso para la identificación de duplicados:

- Creación de un id único por registro que es la unión de los siguientes campos: [año+juzgado+número de acta+libro+nombre(s)+apellido(s)]
- Identificación de los registros duplicados con respecto a ese identificador único.
- Para registros duplicados se aplican los siguientes criterios:
  - Conservar el registro que tenga menor número de campos vacíos o nulos.
  - En caso de que el número de campos vacíos o nulos sea el mismo, se conserva el registro con una llave primaria mayor.<sup>1</sup>

Este proceso garantiza la fiabilidad de la información dentro de la base de datos y reduce la posibilidad de contener errores en la información.

## Anonimización de la información

Para garantizar la protección de datos personales, se conservan únicamente los campos que no hagan posible identificar a un individuo dentro de la base de datos. Es decir, se eliminan los campos de juzgado, número de acta, libro, nombre(s) y apellido(s) con los que se creó el identificador único además de otros campos con información personal.

En este sentido, la base publicada en el Portal de Datos Abiertos incluye los siguientes campos:

1. Año de nacimiento
2. Mes de nacimiento
3. Fecha de nacimiento
4. Sexo
5. País de nacimiento
6. Entidad federativa de nacimiento
7. Municipio o alcaldía de nacimiento
8. Edad de la persona tutora 1
9. Sexo de la persona tutora 1
10. Nacionalidad de la persona tutora 1
11. Edad de la persona tutora 2

---

<sup>1</sup> La llave primaria es un identificador único para cada registro en una base de datos relacional.



12. Sexo de la persona tutora 2
13. Nacionalidad de la persona tutora 2

Con base en lo anterior, cada observación de la base representa un nacimiento registrado en la Ciudad de México y esta observación tiene el grado necesario de disociación con el cuál se garantiza que no se puede identificar a una persona en particular.

### **Codificación de variables de texto**

La base de datos contiene variables codificadas de acuerdo con catálogos de RENAPO. En particular, los campos que se refieren a países de nacionalidad o nacimiento, municipios y alcaldías se encuentran codificados de acuerdo con los catálogos de RENAPO.

Las variables de año y mes de nacimiento se extraen directamente de la columna de fecha de nacimiento.

### **III. PUBLICACIÓN**

Para la publicación de la información se generó un archivo en formato .csv con las siguientes características:

- Encoding: UTF-8
- Formato: .csv
- Frecuencia de actualización: mensual
- Partición de la base: quinquenal