

Desidentificación y Anonimización de archivos de microdatos estadísticos. La experiencia del INE de Uruguay ¹

Federico Segui ²

Palabras clave: anonimización, protección de microdatos, estadísticas oficiales.

La recopilación y difusión de microdatos estadísticos, producto de la realización de censos, encuestas, así como también provienen de registros administrativos, es de gran importancia para el estudio del comportamiento sociodemográfico y económico de la población, así como también para la toma de decisiones en la aplicación de políticas socioeconómicas para generar el desarrollo de un país, y a su vez fomentar las investigaciones sociales.

En pos de satisfacer la demanda de información no se debe perder de vista el garantizar a los informantes la confidencialidad de los datos brindados, ya que las oficinas productoras de estadísticas oficiales deben honrar los compromisos de confidencialidad asumidos con los informantes. El garantizar que se cumplan los derechos de todos los involucrados no es una tarea sencilla, pero mediante el uso de una herramienta como la Anonimización, ya sea de datos tabulares o de los microdatos, se minimiza el riesgo de divulgar información del informante que pueda afectar su privacidad del informante y permite un adecuado uso de los datos.

La Anonimización es una herramienta de control de divulgación de datos estadísticos que busca minimizar tres de los riesgos clave de la publicación de datos estadísticos. Estos riesgos son la singularización, es decir el riesgo a la identificación de una persona a partir de una relación única de ciertas variables; el riesgo de vincular datos de un archivo de microdatos con otros datos del mismo archivo o de otro para identificar a quien pertenece dicha información y finalmente el riesgo de inferencia, definida como la posibilidad de deducir el valor un atributo a partir de otros datos del archivo en cuestión u otro adicional, con una probabilidad significativa.

¹ Trabajo presentado en las XV Jornadas de Investigación de la Facultad de Ciencias Sociales-UdelaR. Montevideo, 14, 15 y 16 de setiembre de 2016)

² Jefe de la Unidad de Gestión de la Calidad del Instituto Nacional de Estadística del Uruguay. correo electrónico: fsegui@ine.gub.uy

En Uruguay desde hace un tiempo se viene aplicando técnicas de Anonimización para determinar el nivel de riesgo de identificación por parte de los usuarios de la información a publicar, tanto en las operaciones realizadas en el Instituto Nacional de Estadística como en otras oficinas productoras de estadísticas del Sistema Estadístico Nacional.

En este sentido es que esta ponencia abordará la importancia de la aplicación de técnicas de desidentificación y Anonimización de los archivos de microdatos publicados por las Oficinas del Sistema Estadístico Nacional.

Las Oficinas Nacionales de Estadística (ONE) publican cada vez más información procurando que sea de mejor calidad y con un mayor nivel de detalle. Por otra parte, los investigadores reclaman, año tras año, que la difusión de la información estadística esté acompañada de archivos de microdatos para un análisis más profundo del fenómeno de estudio. Sin embargo, estos objetivos - información más detallada y archivos de microdatos - entran en conflicto con la obligación que tienen las ONE de proteger la información individual brindada por los informantes.

Preservar la confidencialidad de los datos es una de las principales preocupaciones legales y éticas de las oficinas nacionales de estadística.

En la mayoría de los países existen leyes o reglamentaciones sobre la protección de datos personales y confidencialidad de la información estadística.

Pero no todo se debe a restricciones legales solamente, existen además, implicancias éticas relativas a la confidencialidad de los datos. Así como también, el impacto negativo sobre las tasas de respuesta y veracidad de los datos recogidos, que podrían llegar a obtener las ONE si no se toman medidas adecuadas para asegurar la confidencialidad de la información suministrada por los informantes.

La clave radica en lograr un balance entre la protección de la información estadística para garantizar la confidencialidad de los datos individuales y difundir información con el nivel de desagregación que requieren los usuarios.

Las técnicas de protección de datos estadísticos tienen por objetivo minimizar las posibilidades de los “intrusos” (usuarios que intentan identificar datos individuales mediante técnicas informáticas y estadísticas) de re-identificar información confidencial. Estas técnicas implican, en algunos casos, pérdida de información detallada, lo que significaría que en ciertas ocasiones no se podrán cubrir las necesidades de determinados usuarios, y este dilema es el que deben enfrentar las ONE a la hora de definir la estrategia de protección de datos. Proteger los microdatos en forma excesiva implica perder calidad en la información y, por el otro lado, no tomar ninguna medida para salvaguardar los datos que se difunden generaría gran desconfianza en las ONE, más allá de las implicancias legales que esto conlleva.

Por otro lado, no siempre se podrán utilizar las mismas técnicas de protección de datos en operaciones estadísticas diferentes, para usuarios distintos con requerimientos disímiles. No es lo mismo proteger archivos de microdatos de encuestas sociodemográficas que aquellos relativos a encuestas de actividad económica. Seguramente deberán ser más estrictas las técnicas de protección a aplicar sobre los microdatos que serán difundidos para el usuario común, que sobre los archivos de microdatos para usuarios más calificados como investigadores y analistas, que en general están identificados por la ONE y hasta se les podría exigir la firma de un contrato que asegure la confidencialidad de los microdatos que se le entregan.

Es importante, además, tomar en cuenta que no existen técnicas infalibles de protección de datos, sino que la idea es aplicar varias a la vez, sin excederse en lo que se quiere proteger ya que siempre habrá alguna pérdida de información, basándose en el nivel de riesgo aceptable establecido en la etapa de análisis de riesgos.

Este proceso puede verse como un ciclo de mejora del balance entre protección y pérdida de información, donde en cada iteración del ciclo se intentan ajustar los métodos de protección llegando al nivel de riesgo aceptable con el mínimo de pérdida de información.

Todos estos aspectos deberán ser considerados a la hora de diseñar y planificar los métodos de protección de datos estadísticos a utilizarse por la ONE. Seguramente esto implica la redacción de una política de difusión institucional que asegure que se han tomado todas las medidas necesarias para preservar la confidencialidad de los datos.