Introducción a la anonimización de datos: Técnicas y casos prácticos

Resumen ejecutivo:

Tanto cuando hablamos de datos abiertos, como de datos en general, es crítico [poder garantizar la privacidad de los usuarios y la protección de sus datos personales](https://datos.gob.es/es/documentacion/orientaciones-sobre-la-proteccion-de-datos-en-la-reutilizacion-de-la-informacion-del%20), entendidos como derechos fundamentales. Un aspecto que en ocasiones no recibe especial atención a pesar de las rigurosas normativas existentes, como el [Reglamento General de Protección de Datos (RGPD).](https://eur-lex.europa.eu/eli/reg/2016/679/2016-05-04?locale=es)

**Las soluciones de anonimización de datos están en constante evolución**, siendo un problema especialmente complejo de abordar, dado que no es posible garantizar una anonimización absoluta. En este sentido, la anonimización se gestiona como un **proceso de análisis de riesgos**, donde se busca ofrecer un **equilibrio entre las garantías de privacidad del conjunto anonimizado y la utilidad que tiene** para una tarea concreta, donde el grado de anonimato se puede medir en una escala.

Un aspecto crítico es que **ninguna técnica aplicada de forma aislada es suficiente para aportar garantías frente a los tres tipos de ataques principales:** singularización (identificar a un individuo concreto), vinculación (asociar uno o varios registros de un individuo en uno o varios conjuntos de datos) e inferencia (probabilidad de deducir atributos de un individuo a partir de otros atributos).

En general existe un **alto grado de desconocimiento**, lo cual hace que existan múltiples [fallos comunes](https://www.aepd.es/es/documento/10-malentendidos-anonimizacion.pdf%20) en estos procesos. **El problema más habitual es considerar que el cifrado o la seudonimización son técnicas adecuadas**, cuando en realidad ni siquiera se podrían considerar técnicas de anonimización, ni los conjuntos resultantes estarían exentos de la aplicación del RGPD, dado que siguen siendo considerados datos de carácter personal. Si bien estas técnicas sí **mejoran la protección frente a la vinculabilidad**, por la ofuscación de datos sensibles.

Dado que además **no existe ninguna prescripción oficial** con respecto al uso de ninguna técnica concreta, **lo más recomendable es aplicar una combinación adecuada**, que incluya al menos técnicas de aleatorización, para enmascarar la correlación de valores con individuos concretos, y de generalización, para alterar escalas u órdenes de magnitud**. La aleatorización es muy útil para minimizar los riesgos de vinculabilidad e inferencia**, que al ser **combinada con técnicas de generalización reduce además el riesgo de singularización drásticamente.**

En este informe vamos a explicar los **conceptos clave de un proceso de anonimización de datos**, incluyendo definiciones, principios metodológicos y tipos de riesgos existentes. Finalmente se detallarán las técnicas esenciales que se aplican en la actualidad, presentado ejemplos prácticos de las más relevantes. Cabe destacar que existen técnicas más avanzadas no recogidas en este informe, siendo **un campo con mucha actividad a nivel académico e investigador.**

Puedes leer el informe completo en [este enlace](https://datos.gob.es/sites/default/files/doc/file/introduccion_a_la_anonimizacion_de_datos-_tecnicas_y_casos_practicos.pdf).