

[PAN]Técnicas de anonimidad (Resumen)

La anonimización de los datos puede ser considerada un mecanismo para sanear la información, de forma que la privacidad de los sujetos referenciados en estos pueda ser garantizada por lo tanto:

- La información personalmente identificable debe ser tratada para prevenir su filtración
- Se minimiza el riesgo de filtrado de información cuando se muestran datos al público general, permitiendo el análisis de datos.
- Muchas regulaciones requieren que el uso de este tipo de mecanismos mantengan la información de ciudadanos a salvo.

Atributos personalmente identificables que deben anonimizarse:

- Identificadores (Atributos que identifican de manera única al individuo): DNI, carnet de conducir, fotos, etc...
- Pseudo-Identificadores: Atributos que combinados pueden identificar a un individuo.

Atributos sensibles vs no sensibles:

- Cualquier atributo que pueda ser enlazado al individuo debe ser considerado sensible, aunque depende mucho del contexto.
- Cualquier atributo que no es relevante para el contexto se puede considerar no sensible.

Prácticas típicas:

- Data Masking: Se ocultan o eliminan valores del dataset de forma que los valores originales no pueden ser recuperados. Estas modificaciones pueden ser realizadas mediante cifrado, mezclado, diccionarios de sustitución o reemplazo de caracteres. Puede ser estático, lo que requiere monitorización de la base de datos y su enmascarado completo o dinámico, cuando los datos se enmascaran cuando se realizan las consultas.
- Pseudoanonimización: Reemplaza identificadores personales con pseudónimos o identificadores falsos. Normalmente se mantiene un enlace interno entre los datos originales y los que se muestran, por lo que pueden ser recuperados revirtiendo los datos usando la información apropiada (Que debe ser altamente protegida).
- Generalización: Se reemplazan valores específicos de rangos amplios o categorías manteniendo los datos relativamente utilizables. Suele requerirse una cantidad de datos muy grande para asegurarse de que los grupos sean lo suficientemente ambiguos sin perder utilidad.
- Data Swapping: Se permutan los datos o se mezclan los valores de de una fila dentro de una misma columna
- Data perturbation: Añade ruido a los datos y realiza redondeo, tratando de mantener los datos utilizables para su análisis.
- Datos sintéticos: En vez de publicar datos reales o anonimizados, se crea un dataset sintético basado en los datos originales. Se suelen usar técnicas de machine learnign para generar dichos datos sintéticos mediante el uso de modelos generativos.

From:
<https://www.knoppia.net/> - **Knoppia**

Permanent link:
https://www.knoppia.net/doku.php?id=pan:res_tecnicas_anonimidad&rev=1736289338

Last update: **2025/01/07 22:35**

